

Computergestützte Modellierung von Protein-Ligand-Systemen

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
Dr. rer. nat.

der Fakultät für
Biologie
an der

Universität Duisburg-Essen

vorgelegt von
Karsten Sewczyk

aus Marl / Kreis Recklinghausen

Februar 2015

Die der vorliegenden Arbeit zugrunde liegenden Experimente wurden am Zentrum für Medizinische Biotechnologie (ZMB) in der Abteilung für Bioinformatik der Universität Duisburg-Essen durchgeführt.

1. Gutachter: Prof. Dr. Daniel Hoffmann
2. Gutachter: Prof. Dr. Bettina Siebers
3. Gutachter: Prof. Dr. Peter Bayer

Vorsitzender des Prüfungsausschusses: Prof. Dr. Ulrich Schreiber

Tag der mündlichen Prüfung: 18. Mai 2015

Für Paul und Oskar.

*Das Leben ist einfach, aber wir
bestehen darauf, es kompliziert zu
machen.*

- Konfuzius -

Abstract

Biomolecules can be analyzed with regard to their structure, dynamics, interactions, and function. Three standard methods for the analysis of these molecules are homology modeling, docking, and molecular dynamics (MD) simulation. In this dissertation, these three methods are established, tested and combined with additional techniques, such as electrostatic models, significance analysis, cluster analysis, and variants of the Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) method. I examine with these computational modeling methods three biomolecular systems: First, various glucose dehydrogenase isoenzymes of the hyperthermophilic archaeon *Sulfolobus solfataricus* are studied and their respective substrate specificity are compared. Here, the main method is homology modeling. It is enhanced by electrostatic models and docking in order to obtain a deeper understanding of the interactions between enzymes and ligands. Second, chitinase B of the enterobacterium *Serratia marcescens* is investigated, including its interactions with known inhibitors. The results from an initial docking simulations are refined by subsequent calculations with the MMPBSA method. Third, different SH3 domains are examined in complex with their peptide ligands by MD simulations. The simulations are compared statistically with a significance analysis method and clustering. The outcome of these analyses promise to be more realistic models than models developed by conventional methods.

Zusammenfassung

Biomoleküle können auf Basis ihrer Struktur, ihrer Dynamik oder der von ihnen eingegangenen Wechselwirkungen bzw. Funktion betrachtet werden. Die drei klassischen Verfahren der biomolekularen Modellierung, die für solche Untersuchungen verwendet werden sind die Homologie Modellierung, das Docking und die Molekulardynamiksimulation (MD).

In dieser Promotionsarbeit sollen diese drei Verfahren etabliert, angewendet und mit anderen Verfahren, wie elektrostatischen Modellen, Signifikanzanalysen, Clusteranalysen und Varianten der *Molecular Mechanics Poisson-Boltzmann Surface Area* (MMPBSA) Methode, weiterentwickelt werden, um die nachfolgenden Fragestellungen in drei biomolekularen Systemen zu untersuchen.

Im ersten Fall werden verschiedene Glucose Dehydrogenase Isoenzyme des hyperthermophilen Archaeon *Sulfolobus solfataricus* untersucht und ihre Substratspezifität wird miteinander verglichen. Das Basisverfahren ist hier die Homologie Modellierung. Sie wird durch ein elektrostatisches Modell und ein Docking erweitert, um ein tieferes Verständnis ihrer Wechselwirkungen mit den Substraten zu erhalten.

Im zweiten Fall wird die Chitinase B des Enterobakteriums *Serratia marcescens* mit bekannten Inhibitoren untersucht. Hier bildet das Docking die Grundlage und wird unter anderem mit dem MMPBSA Methode erweitert.

Im letzten Fall werden verschiedene SH3 Domänen mit ihren Peptidliganden mittels MD untersucht. Die Simulationen werden dabei mit Signifikanzanalysen statistisch verglichen und geclustert. Daraus ergeben sich Strukturmodelle, die mit höherer Wahrscheinlichkeit zutreffen als Modelle, die mit gängigen Verfahren erzeugt werden.

Inhaltsverzeichnis

Abstract	i
Zusammenfassung	ii
1 Einleitung	1
1.1 Biomolekulare Modellierung	1
1.2 Biologische Systeme	4
1.2.1 Zentraler Kohlenhydratstoffwechsel des <i>Sulfolobus solfa-</i> <i>taricus</i>	4
1.2.2 Chitotriosidase	6
1.2.3 Src Homologie 3 Domäne	7
1.3 Wissenschaftliche Motivation	10
2 Verfahren der Biomolekulare Modellierung	11
2.1 Vergleichende Modellierung	11
2.2 Molekulares Docking	12
2.3 Molekularmechanik und Kraftfeldfunktionen	14
2.4 Molekulardynamiksimulationen	16
2.5 Energieminimierungen	21
2.6 Trajektorien- und Strukturmodellanalyse	22
2.6.1 Kontinuierliche Solvatationsmodelle	22
2.6.2 MMPBSA Verfahren	25
2.6.3 Fittingmethoden	27
2.6.4 Clusteranalyse	28
2.6.5 Signifikanzanalyse	30
3 Systeme und Anwendungen	33
3.1 Modellierung spezifischer GDH Isoenzyme	33
3.1.1 Einführung	33
3.1.2 Material und Methoden	34
3.1.3 Ergebnisse und Diskussion des Isoenzym SSO_3204 . . .	35
3.1.4 Ergebnisse und Diskussion des Isoenzym SSO_3024 . . .	37

3.2	Modellierung von Inhibitorpositionen in Chitinase B	46
3.2.1	Einführung	46
3.2.2	Materialen und Methoden	48
3.2.3	Ergebnisse und Diskussion	51
3.2.4	<i>Re</i> -Scoring	55
3.3	Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen	58
3.3.1	Einführung	58
3.3.2	Material und Methoden	60
3.3.3	Ergebnisse und Diskussion	63
3.4	Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen	76
3.4.1	Einführung	76
3.4.2	Material und Methoden	81
3.4.3	Ergebnisse und Diskussion	83
Ausblick		107
References		109
Publikation		128
Danksagung		134
Erklärungen		135

Abkürzungsverzeichnis

AASS	Average of Auto-Scaled Scores
ALJ	<i>cyclo</i> -(L-Arg-L-Pro)
AMI	Allosamizolin
CI-4	<i>cyclo</i> -[L-Arg-D-Pro]
DBI	Davies-Bouldin Index
DNA	Desoxyribonukleinsäure
ED	Entner-Doudoroff
EM	Energieminimierung
FDR	false discovery rate
GAD	Gluconat Dehydrogenase
GDH	Glucose Dehydrogenase
GDL	2-Acetamino-2-deoxy-D-glucono-1,5-lacton
GIO	<i>cyclo</i> -(L-Gly-L-Pro)
KDG	2-Keto-3-desoxygluconat
KDPG	2-Keto-3-desoxy-6-phosphgluconat
Lck	Lymphozyten-spezifische Kinase
MC	Monte Carlo
MDR	Mittelkettige Alkohol/Polyol-Dehydrogenase/Reduktasen
MMGBSA	Molecular Mechanics Generalized Born Surface Area
MMPBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
MT	Mutante
NAA	N-Acetyl-D-allosamin
NAD	Nicotinamidadenindinukleotid
NADP	Nicotinamidadenindinukleotidphosphat
NAG	N-Acetyl-D-glucosamin
NGO	2-Methyl-4,5-dihydro-(1,2-dideoxy- α -D-glucopyranoso)[2,1-D]-1,3-oxazol
NMA	Normalmoden Analyse
OHZ	<i>cyclo</i> -(L-Arg-D-Pro)
PHJ	N-[(aminooxy)carbonyl]-N-phenylamin
PME	Particle-Mesh Ewald

PPII	Polyprolin II Helix
RMSD	root mean square deviation
RMSF	root mean square fluctuation
SH3	Src Homologie-3
SAM	significance analysis of microarrays
SulfoSYS	Sulfolobus Systems Biology
TYP	<i>cyclo</i> -(L-Tyr-L-Pro)
WT	Wildtyp
ZKS	Zentraler Kohlenhydratstoffwechsel
3D	dreidimensional

1 Einleitung

1.1 Biomolekulare Modellierung

Die molekulare Modellierung hat einen langen Weg seit ihrer Gründerzeit durchlaufen. Sie hat Einzug gehalten in jeden elementaren Chemieunterricht, um mit Kugel-Stab Modellen zu veranschaulichen, wie Moleküle aus Atomen zusammen gesetzt sind. Diese Modelle werden aber auch in der Wissenschaft und Forschung verwendet [1]. Als eines der bekanntesten Modelle ist in diesem Zusammenhang das DNA (Desoxyribonukleinsäure) Modell von Francis Crick und James Watson im Jahr 1953 zu nennen [2, Abbildung 1.1]. Beide erhielten im Jahre 1962 zusammen mit Maurice Wilkins für diese Entwicklung, dass das Erbmolekül, die Struktur einer Doppelhelix aufweist, den Nobelpreis für Physiologie und Medizin. Heute versteht man unter der molekularen Modellierung die Anwendung von rechnergenerierten Modellen. In molekularen Studien werden Moleküle aus wenigen Atomen bis hin zu umfangreichen Komplexen aus Biomolekülen dargestellt [1]. Bei der Durchführung einer biomolekularen Modellierung kann auf verschiedenen Niveaus gearbeitet werden (Tabelle 1.1). Dabei ist das gewählte Niveau von dem zu beschreibenden biomolekularen Prozess und der Fragestellung abhängig [5]. Die meisten Studien beinhalten drei Schritte. Im ersten Schritt wird ein Niveau ausgewählt, das die intra- und intermolekularen Wechselwirkungen des Systems beschreibt (Tabelle 1.1). Der zweite Schritt der Modellierung beinhaltet die Anwendung des Niveaus auf das System selbst. Im dritten abschließenden Schritt erfolgt eine Analyse der Berechnung, um die Eigenschaften des Systems auszuwerten und um sicherzustellen, dass die Berechnungen zuverlässig sind [6].

In der hier vorliegenden Arbeit wird auf dem Niveau der klassischen statistischen Mechanik, der statistischen Methoden und Kontinuummethoden gearbeitet. Dabei wird die klassische statistische Mechanik vorwiegend in den Kapiteln 3.2, 3.3 und 3.4 eingesetzt, in denen Verfahren wie Molekulardynamiksimulationen (Kapitel 2.4) und Docking (Kapitel 2.2) genutzt werden, um bestimmte Peptide (Kapitel 1.2.2 und 1.2.3) gegen menschliche tödliche Krankheiten zu studieren. Auch wenn beide Verfahren sich in ihrer Genauigkeit und ihrem Ziel unterscheiden, basieren doch beide auf Sampling und Scoring [7]. Mit statistischen Metho-

1.1 Biomolekulare Modellierung

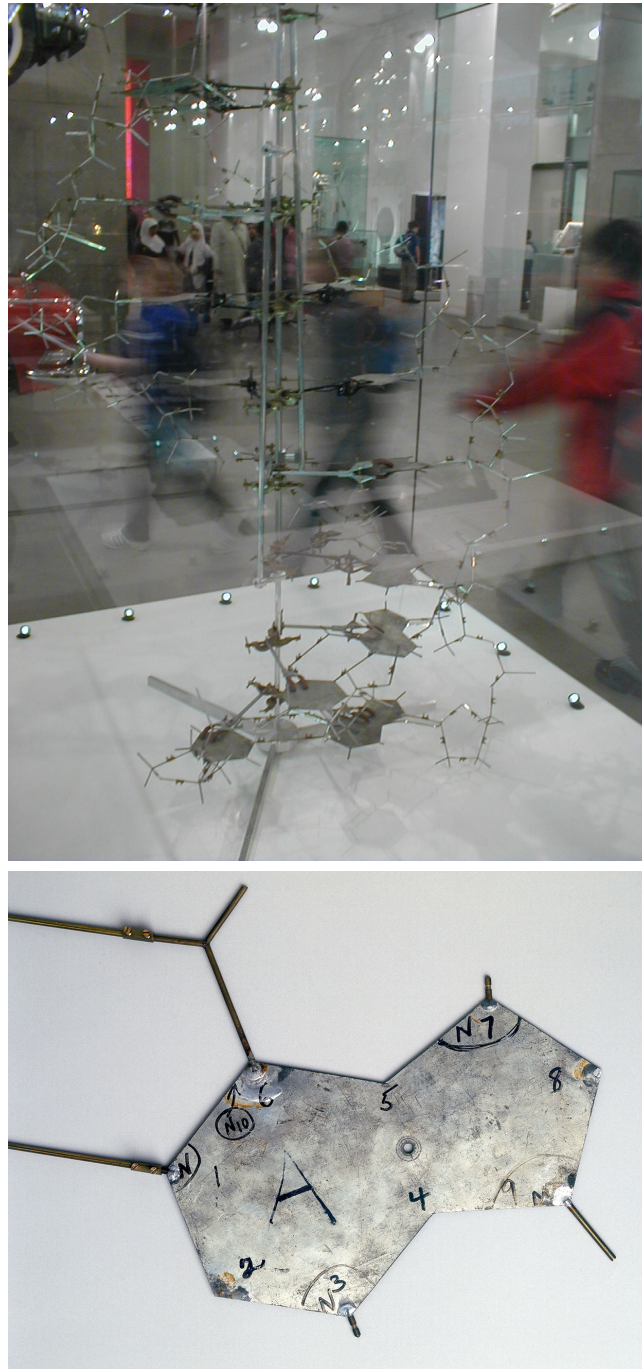


Abbildung 1.1: Original Modell der DNA Doppelhelix. OBEN: Original Modell der DNA Doppelhelix von Crick und Watson, ausgestellt im National Science Museum von London [3]. UNTEN: Aluminium Modell der Base Adenin (A) aus dem DNA Modell von Crick und Watson [4].

1.1 Biomolekulare Modellierung

den und Kontinuummethoden wird vorwiegend in Kapitel 1.2.1 gearbeitet. Hier werden homologe Modelle (Kapitel 2.1) eines enzymatischen Systems (Kapitel 1.2.1) erstellt und mit einem kontinuierlichen elektrostatischen Modell (Kapitel 2.6.1) ausgewertet.

Tabelle 1.1: Niveaus der biomolekularen Modellierung. Klassifizierung der gebräuchlichsten Methoden zur biomolekularen Modellierung. Für die hier vorliegende Arbeit werden Methoden der klassischen statistischen Mechanik sowie statistische Methoden und Kontinuummethoden verwendet [5].

Methoden	Freiheitsgrade	Eigenschaften, Verfahren	Zeitskala
Quanten-dynamik	Atome, Kerne, Elektronen	angeregte Zustände, Entspannung, Reaktions-dynamik	Pikosekunden
Quanten-mechanik	Atome, Kerne, Elektronen	Grund- und angeregte Zustände, Reaktionsme- chanismen	keine Zeitskala
klassische statistische Mechanik	Atome, Lösemittel	Ensembles, Mittelwerte, Systemei- genschaften, Faltung	Nanosekunden
statistische Methoden	Atomgruppen, Aminosäuren, Basen	strukturelle Homologie	keine Zeitskala
Kontinuum- methoden	Elektrisches-, Geschwindigkeits- kontinuum usw.	rheologische Eigenschaften	supramolekular
kinetischen Gleichungen	Populationen von Arten	Populations- dynamik, Signaltrans- duktion	makroskopisch

1.2 Biologische Systeme

1.2.1 Zentraler Kohlenhydratstoffwechsel des *Sulfolobus solfataricus*

Sulfolobus solfataricus ist ein hyperthermophiles Archaeon, welches ein Optimum bei 80°C und pH 3,5 aufweist. Es wurde ursprünglich von einem solfatarischen Feld in Pisciarelli (Italien) isoliert [8] und als Modellorganismus für hyperthermophile Archaeen etabliert. Dazu wurden die Genomsequenz [9], genetische Systeme [10], biochemische Untersuchungen [11], funktionelle Genomik, und ein systembiologischer Ansatz angewendet [12, 13]. In einem Projekt mit dem Namen SulfoSYS (Sulfolobus Systems Biology) [13] wurden auf Systemebene die Auswirkungen von Temperaturänderungen auf den ZKS (zentraler Kohlenhydratstoffwechsel; Abbildung 1.2) erforscht. Genutzt wurden zusätzlich zu enzymatischen Informationen Daten aus der Genomik, Transkriptom, Proteomik und Metabolomik, um ein *in silico* Zellmodell zu bauen [12, 11]. Eine wesentliche Voraussetzung ist dazu die Rekonstruktion und die Identifizierung aller Bestandteile im ZKS Netzwerk, um ausführliche biochemische Information für Enzyme bereitzustellen und das Verhalten und die Regulierung des Netzwerkes zu analysieren. *Sulfolobus solfataricus* benutzt eine modifizierte Form des ED (Entner-Doudoroff) Weges zum Zuckerabbau, ähnlich anderen hyperthermophilen Archaeen, z.B. *Thermoproteus tenax* [14]. Dieser modifizierte ED Weg ist in Abbildung 1.2 gezeigt, seine Reaktionsschritte sind dabei nummeriert (1-13). Er ist geteilt und besteht aus einem nicht-phosphorylierenden Zweig (3,4,5) und einem halbphosphorylierenden Zweig (8,3,9,10,12) [14, 15]. Hier wird, im Unterschied zum klassischen ED Weg, Glucose nicht am Anfang phosphoryliert, sondern über Gluconat und die GAD (Gluconat Dehydrogenase; 2) in KDG (2-Keto-3-desoxygluconat), einem charakteristischen Zwischenprodukt, umgewandelt. KDG wird anschließend durch die KDG Kinase (8) zu KDPG (2-Keto-3-desoxy-6-phosphogluconat) phosphoryliert oder direkt von der KDPG Aldolase (3) gespalten. Im zweiten Fall wird die Phosphorylierung durch den Reaktionsschritt (4), zum Glycerat, durchgeführt. Beide Zweige sind sowohl für Glucose als auch für Galactose nutzbar [15], da die Enzyme der Reaktionsschritte (1A, 2, 8 und 3) Zucker unspezifisch arbeiten [15, 16, 17, 18].

1.2 Biologische Systeme

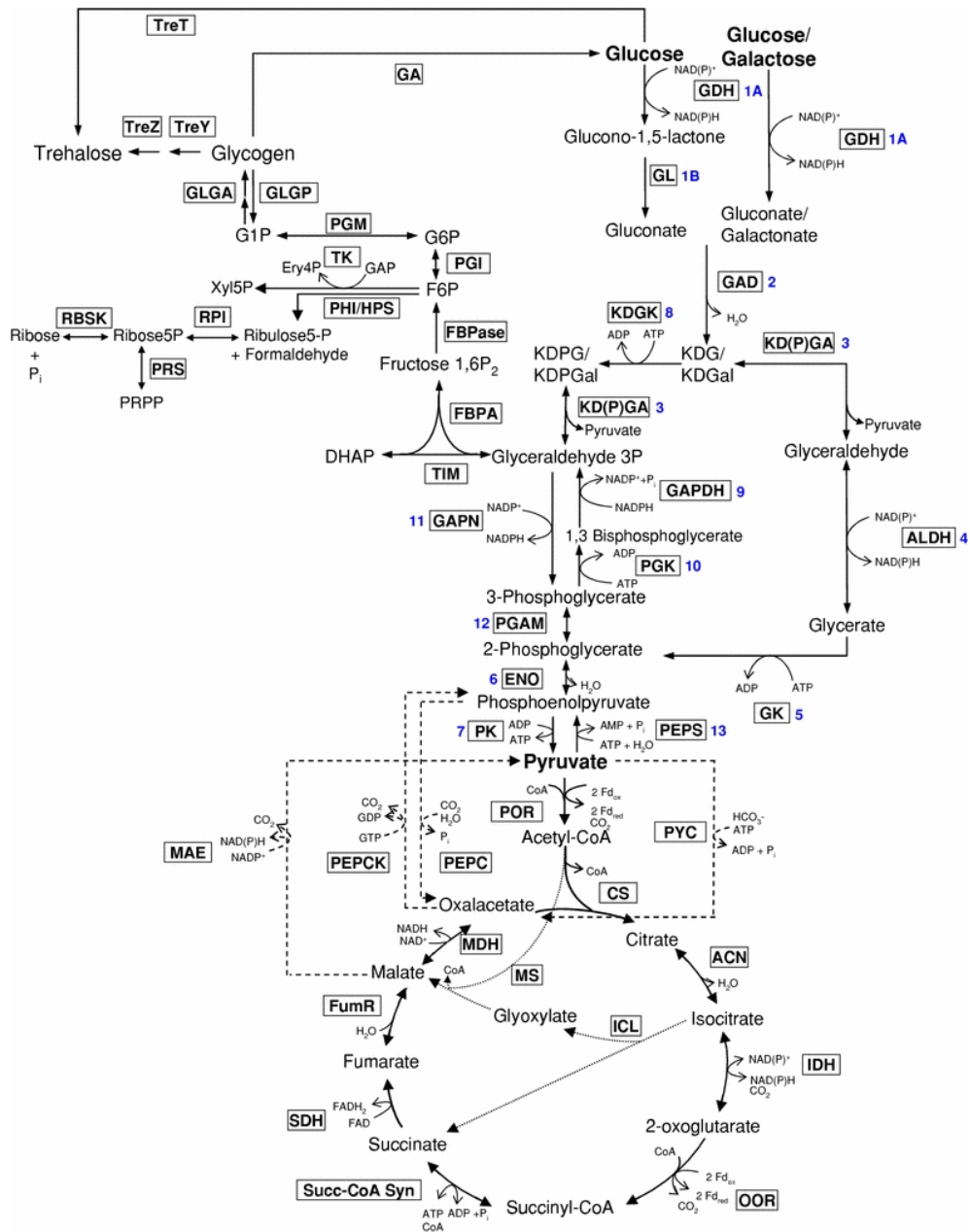


Abbildung 1.2: Zentraler Kohlenhydratstoffwechsel von *Sulfolobus solfataricus*. Identifizierte Reaktionen des zentralen Kohlenhydratstoffwechsels von *Sulfolobus solfataricus*. Gezeigt ist der verzweigte ED Weg mit den Reaktionsschritten 1-13, der Zitronensäurezyklus, der Glyoxylatzzyklus (gepunktete Pfeile), der umgekehrte Ribulosemonophosphat Weg, die C3 / C4Konvertierung (gestrichelte Pfeile) als auch in Form von Glycogen und Trehalose Metabolismus [19].

Das erste Enzym, die GDH (Glucose Dehydrogenase; 1A), katalysiert Glucose zu Gluconat NAD(P)⁺ abhängig. Die Reinigung von Zellextrakten [20] zeigte aber, dass sowohl NAD⁺ als auch NADP⁺ als Cosubstrat verwendet werden können. Darüber hinaus konnte eine breite Substratspezifität gezeigt werden. GDH gehört zur Familie der MDR (Mittelkettige Alkohol/Polyol-Dehydrogenase/Reduktasen), welche zur Superfamilie der pyridin-nucleotidabhängigen Alkohol/Polyol/Zucker-Dehydrogenasen gehören [21]. Eine enzymatische Charakterisierung des rekombinanten Enzyms [15, SSO_3003] sowie eine Kristallstruktur [22] existieren bereits.

1.2.2 Chitotriosidase

Chitotriosidasen (Chitinasen) sind Enzyme, die β -1,4-glucosidische Bindungen in Poly-N-Acetylglucosamin-Ketten (Abbildung 1.3) hydrolysieren. Gefunden werden diese Bindungen in Chitin, dem zweithäufigsten Kohlenhydratpolymer nach Zellulose [23]. Chitinasen sind in jedem lebenden System von Bakterien über Pflanzen bis hin zu Tieren, einschließlich Arthropoden und Säugetieren bekannt. In Abhängigkeit von ihrer Lokalisation sind sie beteiligt an Häutung von Gliederfüßern, Verdauung, Verteidigung/Immunität und Pathogenität [24]. Bakteriellen Chitinasen sind in den Familien 18, 19 und 23 der Glycosylhydrolasen zu finden. Die meisten bakteriellen Chitinasen gehören allerdings der Familie 18 an. Basierend auf ihrer Sequenzähnlichkeit können diese Chitinasen in die drei Unterfamilien A, B und C unterteilt werden. Hierbei gilt es zu beachten, dass diese Unterscheidung unabhängig von dem Namen der Chitinasen ist. Zum Beispiel gehört die Chitinase B des Bakteriums *Serratia marcescens*, welche in dieser Arbeit verwendet wird, zur Unterfamilie A. Die familiäre Klassifizierung basiert hierbei auf ihrer gemeinsamen Sekundärstruktur. Alle Enzyme der Familie 18 beschreiben ein TIM-Barrel, welches durch acht α -Helices und acht parallele β -Faltblätter beschrieben wird. Der katalytische Mechanismus wird als *retaining* bezeichnet, was sich auf die anomere Konfiguration nach der Spaltung bezieht. Protonendonator für diesen Schritt ist ein Glutamatrest und der entsprechende Protonenakzeptor ist die Acetamidogruppe des Substrates [25]. Inhibitoren für Chitinasen wurden schon gegen Insekten und Pilze, Malaria und entzündliche Reaktion, Asth-

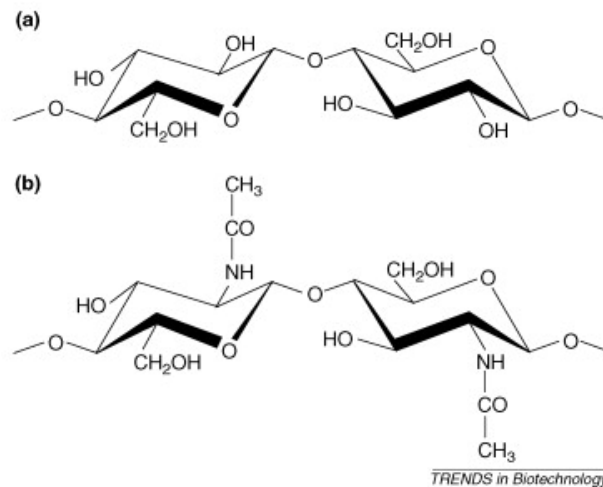


Abbildung 1.3: Chemische Struktur von Zellulose (a) und Chitin (b) [23]. Chitin ist ein polymer, aufgebaut aus Acetylglucosamin Einheiten. Diese Monomere sind durch β -1,4-glycosidische Bindungen miteinander verbunden.

ma, atopische Dermatitis oder allergischem Rhinitis gefunden. Die meisten Inhibitoren gegen die Familie 18 Chitinasen sind Naturprodukte [26]. Prominente Inhibitoren sind hier pseudotrisaccharide wie Allosamidin, Xanthin-Derivate wie C₂-dicafeine und zyklische Peptidderivate, wie CI-4 (*Cyclo*-[L-Arg-D-Pro]) [27]. Zwei cyclische Pentapeptide, genannt Argifin und Argadin, wurden bereits in bestimmten Arten von Pilzen gefunden und hemmen Chitinasen in Insekten. Sie ahmen das natürliche Substrat der Chitinasen nach und haben somit einen Einfluss auf die Häutung von Kakerlakenlarven, was die Substrate zu einem potentiellen Insektizid macht. Humane sowie Chitinasen von anderen Lebewesen werden ebenfalls beeinflusst, aber nicht Chitinase C von *Serratia marcescens*. Dies zeigt die Möglichkeit, Chitinasen unter der Verwendung von Peptiden und deren Derivate selektiv zu hemmen [24].

1.2.3 Src Homologie 3 Domäne

Protein-Protein und Protein-Peptid Interaktionen sind die Grundlage aller zellulären Prozesse. Es existieren zahlreiche Proteindomänen, die solche speziellen Wechselwirkungen vermitteln. Ein prominentes Beispiel ist in diesem Zusammen-

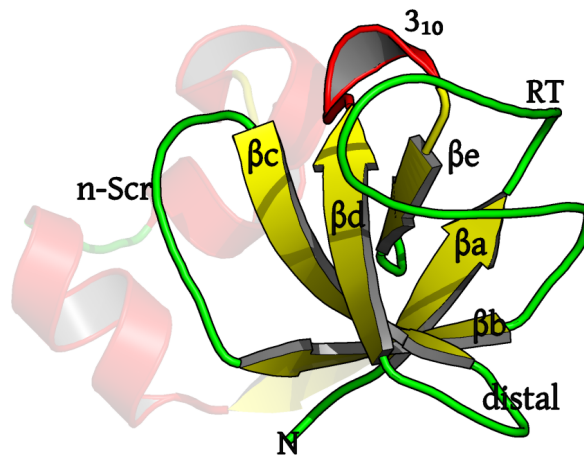


Abbildung 1.4: Sekundärstruktur von SH3 Domänen. Die Sekundärstruktur der Hck-SH3 Domäne ist gezeigt und beschriftet. Vernebelt dargestellt ist außerdem die n-Src-Schleife der PI3K-SH3 Domäne.

hang die SH3 (Src Homologie 3) Domäne [28], von der gezeigt werden konnte, dass sie an Krankheiten wie Krebs, Leukämie, Alzheimer, Osteoporose, entzündlichen und allergischen Prozessen, Asthma und Virusinfektionen (einschließlich AIDS und Hepatitis C) beteiligt ist.

SH3 Domänen haben eine Größe von etwa sechzig Aminosäuren und bestehen aus fünf anti-parallelen β -Faltblättern (Abbildung 1.4). Die ersten vier dieser Faltblätter (β_a , β_b , β_c , β_d) sind durch drei Schleifen verbunden, genannt RT-, n-Src- und distale-Schleife. Eine kurze 3_{10} -Helix verbindet den letzten Strang (β_e) mit den anderen. Vor allem die RT- und n-Src-Schleifen sind verantwortlich für die Spezifität der verschiedenen SH3 Domänen [29, 30].

Die meisten SH3 Domänen erkennen PPII (Polyprolin II Helix) Peptidsequenzen. Das minimale Sequenzmuster dieser Peptide ist PXXP, wobei P Prolin bezeichnet und X für jede andere Aminosäure stehen kann, welches in drei flache hydrophobe Taschen der SH3 Domäne bindet [31]. Das Sequenzmuster kann in zwei Richtungen in die Peptide eingebaut werden, welche damit als pseudo-symmetrische Liganden bezeichnet werden. Wenn dieses Muster, wie in Tabelle 1.2 gezeigt, um zusätzliche Aminosäuren erweitert wird, können die Liganden in zwei Klassen aufgeteilt werden, sogenannte Klasse I und II Liganden bzw. Peptide. Durch diese

Einteilung sind Klasse I Liganden N-terminal mit Aminosäuren erweitert worden und Klasse II Liganden C-terminal [32]. Oft ist eine basische Aminosäure Teil dieser zusätzlichen Aminosäuren. Darüber hinaus ist sie meist an der Ligandenposition P_{-3} zu finden [33, Tabelle 1.2]. Diese Position wird auch Ankerposition genannt und bindet in die sogenannte Komparttasche der SH3 Domänen. Die Wechselwirkungen in dieser Tasche sind stark konserviert und an der Unterseite der Tasche zu finden [34, 35]. Im Jahr 2004 untersuchten Fernandez-Ballester et al. den Zusammenhang von Bindungsmustern in SH3 Domänen und den Sequenzen ihrer Liganden [36]. Sie fanden heraus, dass nicht alle SH3 Domänen, die Peptide der Klasse I binden, auch die gleichen Seitenkettenkonformationen zeigen. Deshalb führten sie eine dritte Klasse, die Klasse I', ein. Eine Unterscheidung zwischen Klasse I und I' ist ihren Untersuchungen nach über die Orientierung der zwei konservierten SH3 Aminosäuren W36 und P51 möglich. Schmidt et al. stellten in einer späteren Arbeit fest, dass eine Analyse der Orientierung von P51 ausreicht, um zwischen Klasse I und Klasse I' Liganden bzw. Peptiden zu unterscheiden [37]. Es wird darauf hingewiesen, dass in dieser Arbeit Aminosäuren in SH3 Domänen nach Larson et al. bezeichnet werden [38].

Tabelle 1.2: Nomenklatur und Bindungsmuster von SH3 Peptide Liganden.

Generelle Nomenklatur von Peptide Liganden die an SH3 Domänen binden. Zusätzlich ist das Bindungsmuster von Klasse I und II Liganden gezeigt. P bezieht sich dabei auf Proline, X auf eine beliebige Aminosäure, und + auf eine positiv geladene Aminosäure. PD1 und PD1R bezieht sich auf Peptide, die in dieser Arbeit verwendet werden. Ihre Sequenzen sind im Ein-Buchstaben-Code beschrieben.

	Position											
	P_{-6}	P_{-5}	P_{-4}	P_{-3}	P_{-2}	P_{-1}	P_0	P_1	P_2	P_3	P_4	P_5
Klasse I				+	X	X	P	X	X	P		
Klasse II						P	X	X	P	X	+	
PD1	H	S	K	Y	P	L	P	P	L	P	S	L
PD1R	H	S	K	R	P	L	P	P	L	P	S	L
	1	2	3	4	5	6	7	8	9	10	11	12

1.3 Wissenschaftliche Motivation

Die wissenschaftliche Literatur basiert oft auf erfolgreichen Geschichten und Happy End Beispielen. Dem vorausgegangen sind aber oft unzählige Rückschläge und misslungene Versuche. Um kritische Hindernisse zu umgehen sind im Forschungsalltag neue und kreative Strategien gefragt [6]. Ein wesentlicher Punkt, welcher mir am Anfang meiner Forschungskarriere in der biomolekularen Modellierung auffiel, war die Auswahl und Analyse von Konformationen in einem mehr-trajektorien Ansatz von Molekulardynamiksimulationen. Bei einer Literaturrecherche erwartete ich einen praktischen Leitfaden für dieses Problem. Stattdessen fand ich zumeist historisch gewachsene oder überlebte Workflows und Best Practices. Ein weiteres Problem der täglichen Forschung ist, dass aus Gründen von Zeit und Ressourcen einige Techniken nur rudimentär verwendet werden, und dass viele interessante und wichtige Anwendungen nicht im Detail beschrieben sind.

Ein Ziel dieser Arbeit ist es, eine neue praktische Anleitung für die Handhabung, Auswahl und Analyse von mehr-trajektorien Ansätzen in Molekulardynamiksimulationen zu entwickeln. Solche Ansätze werden normalerweise, wie auch in diesem Fall, spezifisch für ein bestimmtes biologisches Systemen entwickelt. Es soll ein Ergebnis realisiert werden, welches auf andere biologische Systeme ohne komplizierte Anpassungen übertragen werden kann.

Ein Charakteristikum der biomolekularen Modellierung ist ihr schnelles Wachstum, nicht zuletzt durch die permanente Weiterentwicklung von Computer Hardware und Software in den letzten Jahren. Die größten Projekte des verteilten Rechnens zu Beginn dieser Arbeit sind bei Abschluss dieser Arbeit mit nur einem Desktop Rechner möglich. Zum Beispiel verkürzt die Entwicklung von AutoDock Vina die Rechenzeit eines Systems auf das 62-tel im Vergleich zur Entwicklung mit AutoDock 4 [39]. Eine weitere Entwicklung auf VinaLC [40], die das Programm mit Hilfe eines MPI und Multithreading-Hybridschema parallelisiert, verkürzt eine Berechnung von einem Monat auf eine Stunde [41]. Es ist zu erwarten, dass Techniken, die in dieser Arbeit verwendet und als Stand der Technik vorgestellt werden, in nur wenigen Jahren veraltet sein werden.

2 Verfahren der Biomolekulare Modellierung

2.1 Vergleichende Modellierung

Die vergleichende Modellierung gehört zum Niveau der statistischen Methoden (Tabelle 1.1) und wird auch als Homologie Modellierung oder wissensbasierte Modellierung bezeichnet. Sie ist in Abwesenheit einer experimentell abgeleiteten 3D (dreidimensionalen) Proteinstruktur die nützlichste Technik, um ein Modell einer bestimmten Struktur darzustellen [42, 43, 44, 45]. Die vergleichende Modellierung macht sich die Eigenschaft zu nutze, dass Proteinstrukturen besser konserviert sind als Proteinsequenzen. Somit lassen sich auch für Strukturen mit relativ geringen Sequenzidentitäten hohe strukturelle Ähnlichkeiten finden [43]. Der erste Schritt in der Homologie Modellierung ist es, einen oder mehrere verwandte Proteine mit experimentell kristallisierten Strukturen, sogenannte Templates, zu identifizieren, die in einer Datenbank (z.B. der Protein-Datenbank [46]) gefunden werden können. Mit Programmen wie BLAST oder FASTA [47, 48, 45] können auf einfache Weise Sequenzdatenbanken nach homologen PDB Strukturen zur Zielsequenz gescannt werden. Dabei werden die Strukturen mit den höchsten Sequenzidentitäten als Templates ausgewählt. In der Regel ist eine Identität von mindestens 25 – 30 % notwendig, um ein nützliches Modell [43, 45] erstellen zu können.

Der zweite und wichtigste Schritt in der vergleichenden Modellierung ist die Ausrichtung der Zielsequenz auf die Templatestrukturen. Dieser Schritt ist auch als die Erstellung eines Alignments bekannt. Die Ausrichtung des Alignments ist dabei entscheidend, um ein 3D Modell der Zielsequenz mit möglichst hoher Wahrscheinlichkeit zu erhalten. Ein Ähnlichkeitswert wird dabei durch eines der beiden am häufigsten verwendeten Sequenzalgorithmen, BLAST und FASTA, berechnet. Vor dem Start des Modellierungsprozesses sollte sichergestellt werden, dass sich keine Insertionen und Deletionen innerhalb der Sekundärstrukturelemente befinden. Ist dies der Fall sollte das Alignment überarbeitet werden, so dass sich diese Bereiche in Schleifenregionen befinden. Weitere Informationen über das Protein, beispielsweise darüber, welche Aminosäuren sich im aktiven Zentrum befinden, sollten ebenfalls berücksichtigt werden [42, 43, 44].

Der dritte Schritt in der wissensbasierten Modellierung ist die Erstellung des Zielmodells selbst. Der Kern, die strukturell konservierten Regionen, werden durch Kopieren der Koordinaten aus der Struktur des Templates in das Zielmodell übertragen [45]. Insertionen und Deletionen werden, wie oben angegeben, in die Schleifenregionen eingebaut.

Der letzte Schritt in der Homologie Modellierung ist die Beurteilung des konstruierten Modells. Dabei wird die Richtigkeit der Gesamtstruktur, anhand von stereochemischen Faktoren wie Bindungswinkel, Bindungslänge oder Diederwinkel, überprüft. Programme wie PROCHECK und WHAT IF helfen diese Modellqualitäten zu überprüfen. Wenn die Auswertung der erzeugten Modelle nicht zufriedenstellend ist, muss entweder ein neues Alignment generiert oder Details im Bauprozess überarbeitet werden. Weit verbreitet ist die Verwendung von statistischen Potenzialen zur Bewertung der Modelle. Die Methoden PROSA [49] und DOPE [50] haben sich dabei in der vergleichenden Modellierung etabliert.

2.2 Molekulares Docking

Das molekulare Docking ist ein Verfahren, um Strukturen von intermolekularen Komplexen vorher zu sagen. Es ist weit verbreitet bei der Darstellung von Proteininhibitoren, Medikamenten und Peptiden. Von den meisten Dockingalgorithmen wird eine große Anzahl von möglichen Strukturen generiert, ein sogenanntes Sampling durchgeführt, und eine Bewertungsfunktion angewendet (Scoring), um die wahrscheinlichen Bindungsstellen zu identifizieren. In dieser Hinsicht ist es analog zu Molekulardynamiksimulationen, bei der eine Bewegungsgleichung (Kapitel 2.4) als Samplingverfahren und das Kraftfeld (Kapitel 2.3) als Scoringfunktion fungieren [7]. Leider wird durch die Erzeugung und Bewertung von plausiblen Strukturen auch ein sogenanntes Docking-Problem geschaffen [51]. Es basiert auf den Freiheitsgraden, in welchen ein Körper in drei voneinander unabhängige Richtungen bewegt (Translation) und in drei voneinander unabhängigen Ebenen gedreht (Rotation) werden kann, zusätzlich zu den konformativen Freiheitsgraden für jedes Molekül. Durch die Verwendung einer interaktiven Computergrafik kann dieses Problem vermieden und plausible Strukturen manuell gefunden werden. Dies ist jedoch nur dann wirksam, wenn der zu erwartende Bin-

dungsmodus ähnlich dem bekannten Bindungsmodus eines anderen Liganden ist. Dennoch haben kristallographische Experimente gezeigt, dass abweichende Bindungsarten auch von sehr ähnlichen Liganden eingenommen werden können. Im Gegensatz dazu arbeiten automatische Dockingalgorithmen vorurteilsfrei. Stetig werden neue Algorithmen entwickelt um ein Docking-Problem zu vermeiden.

Am einfachsten ist das Docking vorstellbar, bei einer Annahme von zweier Moleküle als starre Körper, so dass diese nur noch die oben vorgestellten sechs Freiheitsgrade besitzen. Ein solcher Algorithmus wird in dem Programm DOCK verwendet [52]. Er wurde entwickelt, um Moleküle mit einem hohen Grad an komplementären Formen an einer Bindungsstelle zu finden.

Ein flexibles Docking Verfahren betrachtet zusätzlich die Freiheitsgrade der Konformationsänderungen, wobei die meisten dieser Verfahren konformative Änderungen nur für den Liganden und nicht die für den Rezeptor berücksichtigen. Heute angewendete Verfahren berücksichtigen Konformationsänderungen an einem bestimmten Punkt in ihrem Algorithmus. Beispielsweise werden MC (Monte Carlo) Simulationen oft in Kombination mit einer Energieminimierung (Kapitel 2.5) verwendet [53]. Hier werden die inneren Konformationen des Liganden nach jedem MC Schritt geändert oder der Ligand wird zufällig durch die sechs Freiheitsgrade verschoben. Danach wird die Energie mittels eines Kraftfeldes (Kapitel 2.3) berechnet und der Schritt wird, basierend auf voreingestellten Kriterien, angenommen oder abgelehnt.

Genetische Algorithmen können auch für ein Docking verwendet werden [54, 55, 56]. Dabei wird ein sogenanntes Chromosom entweder für eine Liganden-Konformation oder -Orientierung innerhalb des Rezeptors verwendet. Während beide in Abhängigkeit von ihrer laufenden Populationen variieren können. In den nächsten Iterationsschritten werden Individuen anhand ihres Scorings ausgewählt. Des Weiteren können auch Distanzgeometrien fürs Docking verwendet werden, wobei die Entwicklung der Ligand-Konformationen innerhalb der Bindungsstelle das Hauptproblem darstellt. Grenzfunktionen können hier aber modifiziert werden und den Liganden zwingen innerhalb der Bindungsstelle zu bleiben.

Es ist zu beachten, dass das Docking auf vielen Annahmen beruht, die nicht immer jedes Problem bewältigen. Zur Verbesserung bzw. zum korrigieren dieser Abläufe werden deshalb oft zusätzliche Molekulardynamiksimulationen durchge-

führt. Diese sollen helfen, die Flexibilität in die Bindungstasche eines Liganden-Rezeptor-Komplexes genauer abzubilden. Dennoch ist dieses Verfahren teurer als Docking und wird in der Praxis nur für die Veredelung von gedockten Strukturen verwendet. Leider bilden Molekulardynamiksimulationen den Bereich der verschiedenen Bindungsmodi größerer Liganden nicht vollständig ab. Oft liegen höhere Energiebarrieren zwischen diesen Modi, die in solchen Simulationen nur schwer überwunden werden können [6].

2.3 Molekularmechanik und Kraftfeldfunktionen

Wenn wir die Umwelt in unserem täglichen Leben beschreiben, verwenden wir meist die klassische Mechanik. Das traditionelle Beispiel für die klassische Mechanik ist das Heben einer Masse oder das Strecken einer Feder. Dabei werden externe Kraftfelder überwunden, in diesen Beispielen die Schwerkraft oder das Hooksche Gesetz. Die Abbildung komplexer biomolekularer Strukturen ist auch durch Modelle möglich, diese sind hingegen oft anspruchsvoller. Sie müssen für eine große Anzahl von unterschiedlichen Massen bzw. Partikel die Dynamik berechnen. Außerdem enthalten solche Systeme eine Elektronenstruktur, die nicht vernachlässigt werden darf. Es ist sehr schwierig und zeitaufwendig alle Quanteneffekte explizit zu berechnen. Ein allgemeiner Ansatz zur Behebung dieser Probleme besteht darin, die Molekularmechanik zu verwenden. Hierbei werden van-der-Waals Radien und Ladungen verwendet, welche von quantenmechanischen Berechnungen und Experimenten abgeleitet sind. Auch das mechanische Modell einer Feder findet in der Molekularmechanik Anwendung und beschreibt den Gleichgewichtsabstand zweier aneinander gebundener Atome. Werte für solche Abstände werden auch aus Experimenten oder Berechnungen abgeleitet.

Wie einleitend angedeutet, werden Atome eines Systems mit der Newtonschen Mechanik beschrieben. Das Grundprinzip, das dieses möglich macht, ist die Born-Oppenheimer Näherung. Es ermöglicht, die Kernbewegung von der Elektronenbewegung zu trennen. Dies ist die Grundlage heutiger Kraftfelder, wie Amber, Charmm, OPLS und Gromos. Zuletzt genanntes ist ein sogenanntes *united atom* Kraftfeld. Kraftfelder dieser Art enthalten keine unpolaren Wasserstoffatome, sie fusionieren mit ihren zugehörigen Kohlenstoffatomen. In Abhängigkeit vom Sys-

tem ist dies rechnerisch sehr wirtschaftlich. Amber, Charmm und OPLS sind sogenannte *all atom* Kraftfelder. Dies bedeutet, dass jedes Atom in dem System explizit beschrieben ist und keine fusionierten Atome existieren.

Im Allgemeinen besteht ein Kraftfeld aus einer Potentialfunktion und aus Parametern. Die Potentialfunktion ist in allen Kraftfeldern sehr ähnlich. Unterschiede sind hingegen vorwiegend in der Parametrisierung zu finden. Jedes Kraftfeld ist für einen spezifischen Zweck parametrisiert.

Die allgemeine Energiefunktion $V(\vec{r}(t))$ eines Kraftfeldes ist in Gleichung 2.1 dargestellt. Es ist die Ableitung der Positionsvektoren \vec{r} nach der Zeit t . Diese Gleichung kann in physikalische $V^P(\vec{r}(t))$ und nicht physikalische Wechselwirkungen $V^{np}(\vec{r}(t))$ aufgeteilt werden. Der zweite Term enthält sogenannte *restraints*. Sie stellen künstliche Potentiale dar und schränken damit die Bewegungen des Systems ein. Anwendungen für diese Einschränkungen ergeben sich aus experimentellen Studien oder durch unnatürliche Entwicklungen des Systems. Der erste Term stellt eine Summierung von nichtbindenden $V_{nb}(\vec{r}(t))$ und bindenden Wechselwirkungen $V_b(\vec{r}(t))$ (Gleichung 2.2) dar.

$$V(\vec{r}(t)) = V^P(\vec{r}(t)) + V^{np}(\vec{r}(t)) \quad (2.1)$$

$$V^P(\vec{r}(t)) = V_{nb}(\vec{r}(t)) + V_b(\vec{r}(t)) \quad (2.2)$$

Die nichtbindenden Wechselwirkungen werden durch Gleichung 2.3 beschrieben. Sie sind paarweise additiven und enthalten einen Abstoßungs-, Dispersions- und Coulomb-Term. Die Bedingungen der Abstoßung und Dispersion werden im Lennard-Jones Potential $V_{LJ}(\vec{r}_{ij}(t))$ kombiniert. Geladen oder teilweise geladene Atome wirken durch den Coulomb-Term $V_C(\vec{r}_{ij}(t))$.

$$V_{nn}(\vec{r}(t)) = \sum_{i < j} [V_{LJ}(\vec{r}_{ij}(t)) + V_C(\vec{r}_{ij}(t))] \quad (2.3)$$

Die gebundenen Wechselwirkungen $V_b(\vec{r}(t))$ der Gleichung 2.2 bestehen aus einem Zwei-Körper-, einem Drei-Körper- und einem Vier-Körper-Term (Gleichung 2.4). Der Zwei-Körper-Term $V_B(\vec{r}_{ij}(t))$ beschreibt die Bindungsenergie zwischen den zwei Atomen i und j , der Drei-Körper-Term $V_W(\vec{r}_{ijk}(t))$ die Winkelenergie

2.4 Molekulardynamiksimulationen

zwischen den drei Atomen i , j und k und der Vier-Körper-Term $V_D(\vec{r}_{ijkl}(t))$ die Flächenwinkelenergie zwischen den vier Atomen i , j , k und l .

$$V_b(\vec{r}(t)) = \sum_{i < j} [V_B(\vec{r}_{ij}(t)) + V_W(\vec{r}_{ijk}(t)) + V_D(\vec{r}_{ijkl}(t))] \quad (2.4)$$

Die Terme der beiden Gleichungen 2.3 und 2.4 ergeben sich aus den nachfolgenden Gleichungen 2.5 - 2.9.

$$V_B(\vec{r}(t)) = \sum_{\text{Bindung}} K_d (d - d_0)^2 \quad (2.5)$$

$$V_W(\vec{r}(t)) = \sum_{\text{Winkel}} K_\theta (\theta - \theta_0)^2 \quad (2.6)$$

$$V_d(\vec{r}(t)) = \sum_{\text{Flächenwinkel}} K_\chi (1 + \cos(n\chi - \delta)) \quad (2.7)$$

$$V_{LJ}(\vec{r}(t)) = \sum_{\text{Lennard-Jones}} \epsilon_{ij} ((C_{ij}/r_{ij})^{12} - (C_{ij}/r_{ij})^6) \quad (2.8)$$

$$V(\vec{r}(t)) = \sum_{\text{Coulomb}} q_i q_j / \epsilon_r r_{ij} \quad (2.9)$$

Dabei setzen sich die Kraftfeldparameter dieser Gleichungen für jedes Atom i aus Federkonstanten K , Referenz-Bindungslängen d_0 und Winkel θ_0 , Diederwinkel χ , Multiplizitäten n und Phasen δ , Lennard-Jones Parameter ϵ , C und Teilladungen q zusammen.

Die Tatsache, dass alle Terme der Gleichungen 2.1 - 2.4 additiv sind, basiert auf der Annahme, dass sie unabhängig voneinander sind. Diese Annahme wird eingeführt, um den Rechenaufwand zu reduzieren. Das führt jedoch zu einem Verlust an Genauigkeit, da die Ableitungen in Abhängigkeit von den relativen Atompositionen die atomaren Kräfte und Kraftgradienten beschreiben.

2.4 Molekulardynamiksimulationen

An dieser Stelle muss darauf hingewiesen werden, dass ein Kraftfeld nur die potentielle Energie eines Systems beschreibt. Diese Kräfte werden durch Gleichung

2.4 Molekulardynamiksimulationen

2.10 beschrieben. Um die Bewegungsenergien zu betrachten, muss die Gleichung 2.11 für das jeweilige System gelöst werden.

$$F_i = -\frac{dV}{dr_i} \quad (2.10)$$

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} \quad (2.11)$$

Durch die Komplexität der zahlreichen Atome des Systems kann die Bewegungsgleichung nicht analytisch gelöst werden, sondern schrittweise numerisch. Dieses Vorgehen wird Molekulardynamiksimulation genannt. Es wurden zahlreiche Algorithmen für diese Vorgehen entwickelt. Ein prominentes Beispiel ist der *leapfrog* Algorithmus. Er ist reversibel. Das heißt, es ist möglich, eine Simulation exakt rückwärts laufen zu lassen. Es werden die Geschwindigkeiten v und Orte r des Systems zu unterschiedlichen Zeitpunkten t bestimmt. Dazu zeigen die beiden folgenden Gleichungen (2.12 und 2.13) die mathematischen Grundlagen, wobei $\frac{F(t)}{m}$ die Kraft zu einem bestimmten Zeitpunkt t dividiert durch die auf sie wirkende Masse m ist.

$$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + \frac{F(t)}{m}\Delta t \quad (2.12)$$

$$r(t + \Delta t) = r(t) + v\left(t + \frac{\Delta t}{2}\right)\Delta t \quad (2.13)$$

Neben einem Algorithmus zur Beschreibung eines Systems unter biologischen Bedingungen müssen noch weitere Bedingungen definiert werden. Nachfolgend sind hier die wichtigsten Randbedingungen beschrieben.

Simulierte Boxen und periodischen Randbedingungen In der Praxis ist es üblich, eine Box um die zu simulierende Struktur zu definieren. Um Randeffekte an den Wänden der Box zu vermeiden, werden periodische Randbedingungen eingeführt [57].

Diese sind erforderlich, da die zu simulierende Struktur in der Box translatiert. Versucht ein Teil der Struktur die Box zu verlassen, tritt dieser Teil auf der gegenüberliegenden Seite wieder in die Box ein. Formal werden hierdurch die Grenzen

der Box aufgehoben. Minimale Bildkonventionen sorgen ferner dafür, dass jedes Atom jedes andere Atom nur einmal sieht [58]. Dies funktioniert nur für nicht bindende Wechselwirkungen auf kurze Distanz gut. Bei Wechselwirkungen mit größeren Abständen, wie elektrostatische Wechselwirkungen, funktioniert diese Methode nur eingeschränkt.

Die meisten Simulationsprogramme können mit verschiedenen Typen von Boxen, wie kubische, trikline, dodekaedrische und oktaedrische Boxen arbeiten. In dieser Arbeit werden ausschließlich dodekaedrische Boxen angewendet, da sie durch ihre sechsseitige Doppelpyramide ein relativ kleines Volumen besitzen. Ein kleines Boxvolumen hat den Vorteil, dass für explizite Wassermoleküle (siehe unten), eine relativ geringe Rechenzeit benötigt wird. Bei Simulationen mit expliziten Wassermolekülen wird der Teil der Box, in dem ein Vakuum während der Simulation herrschen würde, mit Wassermolekülen aufgefüllt.

Algorithmen zur Temperaturkoppelung Es gibt viele Ursachen für Temperaturschwankungen im simulierten System. Ein Beispiel sind Reibungsverluste. Sie können in einer Simulation auftreten und für einen Anstieg der Temperatur sorgen. Da dies die Simulationsergebnisse beeinflusst, werden oft Algorithmen zur Temperaturkoppelung verwendet. Das Simulationsprogramm Amber besitzt zwei Modelle, das von Berendsen [59] und das von Nosé und Hoover [60, 61].

Algorithmen zur Einschränkung von Schwingungen Die Geschwindigkeit von Molekulardynamiksimulationen wird maßgeblich von schnell oszillierenden Bindungen zu Wasserstoffatomen im System beeinflusst. Aus diesem Grund wurden Algorithmen wie SHAKE [62] oder LINCS [63] entwickelt. Sie bieten die Möglichkeit von stabilen und schnellen Simulationen. Der LINCS Algorithmus setzt dabei alle im Voraus festgelegten Bindungen nach jedem Simulationsschritt zurück auf ihre ursprüngliche Länge. Dies läuft schneller und stabiler ab als der SHAKE Algorithmus, bei dem die vordefinierte Länge konstant gehalten wird. Für den Fall von Bindungen mit Wasserstoffatomen in Wassermolekülen wurde der SETTLE Algorithmus [64] entwickelt. In dieser Arbeit werden die Algorithmen LINCS und SETTLE verwendet, SETTLE für Wassermoleküle, LINCS für alle weiteren Moleküle.

Elektrostatik in Simulationen Bei Wechselwirkungen auf großer Entfernung sind elektrostatische Wechselwirkungen entscheidend für die strukturelle Stabilität. Normalerweise wird ein Großteil der Rechenzeit für die Berechnung dieser Energien aufgebracht [58]. In den frühen Tagen der Molekulardynamiksimulationen wurde eine Toleranzgrenze gesetzt. Dies bedeutete, dass elektrostatische Wechselwirkungen nur bis zu einer bestimmten Distanz berechnet wurden. Durch diesen Ansatz wurden oft instabile Simulationen beobachtet. Mittlerweile können rasterbasierte Verfahren verwendet werden. Ein Beispiel hierfür ist der PME Algorithmus (Particle-Mesh Ewald) [65, 66].

Mehrere unabhängige Simulationen Heute existieren viele Arten der Molekulardynamiksimulationen (AMD, REMD, FEGS, etc.). Sie wurden alle entwickelt, um das Sampling von Simulationen zu erhöhen und sind oft nicht mehr mit dem oben beschriebenen klassischen Ansatz vergleichbar. Ein Weg zur Erweiterung des Konformationsraumes von klassischen Molekulardynamiksimulationen ist die Variation der Startbedingungen. Solche Variationen werden dieser Arbeit verwendet. Ausführliche Erläuterungen und Untersuchungen zu diesen Variationen sind von Genheden und Ryde geschrieben worden [67] und sind nachfolgend aufgeführt:

VIIIT Velocity-induced independent trajectories

(Geschwindigkeit induzierte unabhängige Simulationen)

SIIT Solvation-induced independent trajectories

(Solvatisierung induzierte unabhängigen Simulationen)

CRPIIT Conformation, rotation, and protonation-induced independent trajectories

(Konformations-, Rotations-, and Protonierungs-induzierte unabhängigen Simulationen)

ACIIT Alternative-conformation-induced independent trajectories

(Alternative-Konformation induzierte unabhängigen Simulationen)

Kombinationen dieser Variationen sind ebenfalls anwendbar und erweitern das Sampling darüber hinaus.

Darstellung von Wasser Unter biologischen Bedingungen ist ein zelluläres System mit Wasser gefüllt. In der Molekularmechanik gibt es eine Vielzahl von Möglichkeiten, Wasser zu beschreiben. Die hier entwickelten Modelle können eingeteilt werden in explizite und implizite Wassermodelle. In Simulationen mit expliziten Wassermodellen wird ein großer Teil der Rechenzeit darauf verwendet Wechselwirkungen von Wassermolekülen zu berechnen. Dennoch ist es oft wichtig, sie explizit zu berechnen, da sie eine entscheidende Rolle bei der Darstellung von elektrostatischen Effekten spielen [58, 57].

Explizite Wassermodelle In Zellen sind die Wechselwirkungen zwischen Lösemittel und gelöster Substanz von besonderem Interesse. In manchen Fällen sind die Lösemittelmoleküle sogar Teil der gelösten Substanz, denn sie sind gebunden und haben die Funktion einer Brücke. Aus diesem Grund werden explizite Lösemittel häufig verwendet. Klassifiziert werden können sie durch die Anzahl der Wechselwirkungspunkte, ihre Flexibilität bzw. Steifheit und bei manchen Modellen durch ihre Polarisierbarkeit. Im Hinblick auf diesen Umfang werden im Folgenden nur die am häufigsten verwendeten Modelle beschrieben.

Ein Drei-Punkt-Modell enthält drei Ladungen, eine auf jedem Atom. TIP3P [68] und SPC [69] sind die beiden bekanntesten Modelle dieser Klasse. Das TIP3P Modell wurde in dieser Arbeit verwendet.

Ein Vier-Punkte-Modell verfügt über ein Dummy-Atom, ausschließlich mit negativer Ladung. Angeordnet ist es in der Nähe des Sauerstoff Atoms, in der Mitte des HOH Winkels. Dieses Dummy-Atom verbessert die Elektrostatik in der Nähe des Wassermoleküls. TIP4P-Ewald [70] ist eine Modifikation des ursprünglichen TIP4P [71] Modells. Diese Modifikation wurde entwickelt um besser mit periodischen Randbedingungen umgehen zu können.

Implizite Wassermodelle Da biologische Prozesse in wässrigem Milieu stattfinden, ist es erforderlich Lösungsmittelleffekte zu berücksichtigen. Dabei ist es oft nicht notwendig, jedes einzelne Wassermolekül im System zu simulieren. In diesen Fällen ist es üblich, implizite Lösemittelmodelle zu benutzen [72]. Hier gibt es unterschiedliche Modelle, die jeweils zu einem speziellen Zweck entwickelt wurden.

In dieser Arbeit werden implizite Wassermodele nicht für Molekulardynamiksimulationen, sondern sie für die Trajektorien- und Strukturmodell-Analyse verwendet. Aus diesem Grund werden sie erst in Kapitel 2.6.1 als kontinuierliche Solvationsmodelle vorgestellt.

2.5 Energieminimierungen

Auf den ersten Blick ähneln sich Energieminimierungen und Molekulardynamiksimulationen sehr. Auf den zweiten Blick fällt aber auf, dass die Motivationen völlig unterschiedlich sind. Die Trajektorie einer klassischen Molekulardynamiksimulation beschreibt das Verhalten eines Systems, physikalisch repräsentativ, über die Zeit. Zum Beispiel können *induced-fit* Effekte durch Molekulardynamiksimulationen beschrieben werden. Die Trajektorie einer EM (Energieminimierung) ist hingegen bedeutungslos, da der Algorithmus versucht, das System in Richtung einer niedrigeren potentiellen Energie und einer Minimierung der in dem System wirkenden Kräfte zu treiben. Die Trajektorie ist somit gelenkt und damit ist nur die Energie am Ende der Trajektorie interessant. Die potentielle Energie eines Systems beschreibt viele lokale Minima, aber nur ein globales Minimum. In der Praxis ist es oft nicht möglich, das globale Minimum für aufwendige Systeme zu finden. Heutzutage existieren eine Menge von Algorithmen, um die Energieminima im System zu finden. Die meisten Methoden umfassen Derivate von verschiedenen Ordnungen. Bekannte Beispiele erster Ordnung ist der *conjugate gradient* und der *steepest descent* Algorithmus, wohingegen der Newton-Raphson Algorithmus repräsentativ für Algorithmen zweiter Ordnung steht. In der vorliegenden Arbeit wird lediglich der *steepest descent* Algorithmus angewendet.

Steepest descent nutzt den Gradienten der potentiellen Energielandschaft. Dies entspricht den Kräften des molekularen Systems und sucht somit einen Weg ins nächste Energieminimum. Durch Bewegung der Atome als Reaktion auf die auf sie wirkenden Kräfte führt dies zur Verringerung der potentiellen Energie.

Formal ist der Kraftvektor definiert als

$$F(r) = -\frac{d}{dr}V(r) \quad (2.14)$$

wobei r der Vektor der Atomkoordinaten ist.

Der Algorithmus ist sehr ineffizient für mehrdimensionale Probleme mit unregelmäßigen Potentialoberflächen und mehreren lokalen Minima. Er arbeitet hingegen sehr effizient bei der Lokalisierung des nächsten lokalen Minimums. Folglich sind globale Bewegungen erforderlich, um das globale Energieminimum zu finden. Außerdem ist der Algorithmus sehr effektiv bei der Beseitigung sterischer Konflikte und entspannende Bindungslängen und Winkel zu ihren kanonischen Werten [58].

2.6 Trajektorien- und Strukturmodellanalyse

2.6.1 Kontinuierliche Solvationsmodelle

In der Praxis werden kontinuierliche Solvationsmodelle (Kontinuummethode) am häufigsten verwendet, um den Solvatisierungsterm der freien Energie ΔG_{sol} berechnen zu können. Sie beschreiben die Änderung der freien Energie bei dem Transfer eines Moleküls vom Vakuum ins Lösungsmittel. Diese Energie kann in drei Komponenten aufgeteilt werden:

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{vdw} + \Delta G_{cav} = \Delta G_{pol} + \Delta G_{apol} \quad (2.15)$$

ΔG_{elec} , ist die elektrostatische Komponente der Gleichung und beschreibt die Energie für polare und geladene gelöste Stoffe aufgrund der Polarisierung des Lösemittels. Es wird als ein gleichförmiges Medium mit einer dielektrischen Konstante ϵ modelliert. ΔG_{vdw} sind die van-der-Waals Wechselwirkungen zwischen dem gelösten Stoff und dem Lösemittel. Der Energiebeitrag ΔG_{cav} berücksichtigt die Bildung eines Hohlraumes des gelösten Stoffes, innerhalb des Lösemittels. Diese Komponente ist ein entropischer Aufwand, der durch die Reorganisation der Lösemittelmoleküle um den gelösten Stoff beeinflusst wird. Es wird Arbeit zum Erstellen des Hohlraums, gegen den Lösemitteldruck, geleistet. Zusätzlich kann in Gleichung 2.15 noch ein Term explizit für Wasserstoffbrücken ΔG_{hb} eingefügt werden. Dies ist sinnvoll bei lokalisierten Wasserstoffbrücken zwischen dem gelösten Stoff und dem Lösemittel.

Für die meisten der heutigen Modelle wird die Berechnung von ΔG_{sol} in eine po-

lare $\Delta G_{pol}(\Delta G_{elec})$ und eine apolare $\Delta G_{apol}(\Delta G_{vdw} + \Delta G_{cav})$ Rechnung aufgeteilt. Aus diesem Grund ist die folgende Diskussion in diese beiden Begriffen unterteilt.

Polarer Solvationsterm Die Poisson-Gleichung ist die Grundlage für jedes dielektrische Kontinuum. Ein Boltzmann-Faktor kann hinzugefügt werden, um die Ionenstärke des Systems zu berücksichtigen. Die Gleichung wird dann Poisson-Boltzmann Gleichung genannt und zeigt folgenden Zusammenhang [73]:

$$\nabla[\varepsilon(r)\nabla\phi(r)] - \varepsilon(r)\kappa(r)^2\sinh[\phi(r)] + 4\pi\rho^f(r)/kT = 0 \quad (2.16)$$

Hier ist $\phi(r)$ das elektrostatische Potential, in den Einheiten von kT/q (k = Boltzmann-Konstante, T = Temperatur in Kelvin, q = Protonenladung), ε ist die Dielektrizitätskonstante und ρ^f die Ladungsdichte. Der Ausdruck κ^2 kann auch durch $1/\lambda^2$ oder $8\pi q^2 I/ekT$ beschrieben werden. λ ist dabei die Debye-Länge und I die Ionenstärke des Lösemittels. Die Variablen ϕ , ε , κ und ρ sind Funktionen des Ortsvektors r .

Auch wenn die Ionenstärke häufig in biomolekularen Berechnungen vernachlässigt wird, bezeichnet man die Ansätze als Poisson-Boltzmann Methoden [74].

Es gibt verschiedene Ansätze zur Lösung dieser Gleichung. Dazu werden die Ladungen und die dielektrische Konstante auf ein Gitter übertragen [75]. Um eine Vakuum- und eine Wasserumgebung darzustellen, wird die Berechnung meist mit zwei unterschiedlichen Dielektrizitätskonstanten, 1 und 80, durchgeführt. Derzeit ist das numerische Lösen der Poisson-Boltzmann Gleichung eine sehr zeitaufwendige Rechnung. Es gibt einfachere implizite Lösungsmodelle, wie die verallgemeinerte Born-Annäherung, auch *Generalised Born* oder GB Rechnungen genannt. Für Annäherungen sind verschiedene GB Verfahren entwickelt worden, wobei ein Molekül als eine Reihe von Kugeln mit Ladungen und Radien angenommen wird [76].

In diesem Modell wird der elektrostatische Beitrag zur freien Energie der Solvation durch folgende Gleichung ermittelt:

$$G_{pol} = 1/2((1/\varepsilon_{solv}) - 1) \sum \sum (q_i q_j) / f_{GB} \quad (2.17)$$

Darin sind q_i und q_j Teilladungen, ε_{solv} ist die Dielektrizitätskonstante des Löse-

mittels. Die Funktion f_{GB} ermittelt den wirksamen Bornradius R_i . Dabei wird r_{ij} ermittelt aus der Interpolation aus einem kleinen und einem großen Abstand der beiden Atomen i und j . Die Lösung für f_{GB} macht den Hauptunterschied zwischen den einzelnen GB Ansätzen aus [77].

Die meisten Programme arbeiten mit verschiedenen Ansätzen zur Berechnung dieser Funktionen und der Radien. Beispielsweise enthält Amber die Algorithmen GB^{HCT} [78], D^{OBC} 1 und 2 [79] und GBN [80]. Kleine arzneistoffähnliche Moleküle werden oft genauer mit diesem Näherungsverfahren berechnet, als mit der Poisson-Boltzmann Berechnungen [81]. Diese Beobachtung wird der numerischen Ungenauigkeit im Lösen von Poisson-Boltzmann Berechnungen geschuldet. Aus diesem Grund ersetzen GB Modelle sehr oft das Poisson-Boltzmann Modell in MMPBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) Rechnungen (Abschnitt 2.6.2), wobei dann von MMGBSA (Molecular Mechanics Generalized Born Surface Area) Rechnungen gesprochen wird.

Apolarer Solvationsterm Kontinuummodelle beschreiben neben polaren auch apolare Effekte. Im Wesentlichen ist dies in den van-der-Waals Kontakten und den entropischen Kosten zur Schaffung eines Hohlraums im Lösemittel verankert (Gleichung 2.15).

Neben der Beschreibung der apolare Solvatationsenergie ΔG_{apol} wird außerdem eine Korrelation zwischen koordinierten Wassermolekülen um den gelösten Stoff und ΔG_{apol} für apolare Moleküle gefunden [82]. Im Falle der Alkane korreliert diese Energie linear mit der Oberfläche der Alkane. Dies ist die Grundlage des SASA Modells [83], welches ein sehr beliebter Ansatz zur Berechnung der apolaren Solvatationsenergie ist. Ein Ansatz dieses Modells für größere Moleküle brauchbar zu machen [84, 85], führt zu:

$$G_{apol} = y_{sasa}SASA + b_{sasa} \quad (2.18)$$

Diese Formel enthält die empirischen Parameter, y_{sasa} und b_{sasa} , die auf Kohlenwasserstoffverbindungen und deren experimentellen freien Solvatationsenergien basieren [83].

Unterschiedliche Parametersätze wurden entwickelt, die manchmal Atomspezi-

fisch sind [86]. Andere Methoden, wie das Hohlraum-Dispersions Modell [84] und das polarisierbare Kontinuummodell [87, 88], wurden ebenfalls entwickelt und sind komplizierter als die SASA Methode. Genheden et al. [89, 90] bewertete die Fähigkeit dieser drei Methoden, um Änderung bei Ligandenbindung vorherzusagen. Er stellte fest, dass keine der Methoden genaue Vorhersagen über eine Vielzahl von verschiedenen Systemen liefert. Im Durchschnitt führt die SASA Methode zu den zuverlässigsten Ergebnissen. Die beiden anderen Verfahren haben die Tendenz, ihre Ergebnisse zu überschätzen.

2.6.2 MMPBSA Verfahren

Das MMPBSA Verfahren wird in der Regel verwendet, um Molekulardynamiksimulationen zu evaluieren und führt zur Bestimmung der freien Bindungsenergie ΔG_{Bind} von einem Enzym-Ligand-Komplex [58]. ΔG_{Bind} wird durch die Gleichung 2.19, mit den Komponenten Komplex ΔG_{AB} , Rezeptor ΔG_A und Ligand ΔG_B , definiert.

$$\Delta G_{Bind} = \Delta G_{AB} - \Delta G_A - \Delta G_B \quad (2.19)$$

Die freie Energie der einzelnen Komponenten ΔG_x ($x = AB, A, B$) setzt sich aus den drei Termen molekulare mechanische Energie ΔE_{MM} , Solvation ΔG_{PBSA} und entropischer Beitrag ΔS zusammen. Gleichung 2.20 beschreibt diesen Zusammenhang, mit T als die Temperatur in Kelvin.

$$\Delta G_x = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (2.20)$$

ΔE_{MM} wird mit Gleichung 2.21 näher beschrieben. Die Fußnoten der Energieterme sind Abkürzungen für die drei inneren Energien Bindung, Winkel und Torsion sowie die Energien basierend auf den elektrostatischen und van-der-Waals Kräften.

$$\Delta E_{MM} = \Delta E_{Bindung} + \Delta E_{Winkel} + \Delta E_{Torsion} + \Delta E_{vdw} + \Delta E_{elec} \quad (2.21)$$

Der in Gleichung 2.20 beschriebene Solvationsterm ΔG_{sol} wird näher durch Gleichung 2.22 definiert. In Abhängigkeit des Lösungsweges für ΔG_{pol} in dieser

Gleichung, wird das MMPBSA Verfahren auch oft MMGBSA Verfahren genannt. Die Grundlagen hierfür sind in Kapitel 2.6.1 zu finden.

$$\Delta G_{sol} = \Delta G_{pol} + \Delta G_{apol} \quad (2.22)$$

Des Weiteren ist in Gleichung 2.20 die Konformationsentropie $-T\Delta S$ als Teil der freien Enthalpie beschrieben. Sie setzt sich additiv aus Translations-, Rotations-, und Vibrationsentropie zusammen. Translations- und Rotationsentropie werden dabei schon implizit durch die statistische Mechanik ermittelt. Offen bleibt die Vibrationsentropie, welche explizit berechnet werden muss [91].

Ein klassischer Weg hierzu ist die Berechnung der Normalmoden des Systems [92, 93, 94]. Das Grundkonzept ist dabei, eine Hessematrix zu lösen. Diese ist die zweite partielle Ableitung der Potentialfunktion des Systems. Eine Hessematrix D ist eine $3N \times 3N$ Matrix und ist in Gleichung 2.23 dargestellt.

$$H(r) = \left[\frac{\partial^2 V(r)}{\partial r_i \partial r_j} \right] = \begin{bmatrix} \frac{\partial^2 V(r)}{\partial x_1 \partial x_1} & \frac{\partial^2 V(r)}{\partial x_1 \partial y_1} & \cdots & \frac{\partial^2 V(r)}{\partial x_1 \partial z_n} \\ \frac{\partial^2 V(r)}{\partial y_1 \partial x_1} & \frac{\partial^2 V(r)}{\partial y_1 \partial y_1} & \cdots & \frac{\partial^2 V(r)}{\partial y_1 \partial z_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 V(r)}{\partial z_n \partial x_1} & \frac{\partial^2 V(r)}{\partial z_n \partial y_1} & \cdots & \frac{\partial^2 V(r)}{\partial z_n \partial z_n} \end{bmatrix} \quad (2.23)$$

Die NMA (Normalmoden Analyse) schätzt die Konformationsentropie durch Messen der Breite der Potential-Energie-Senke eines jeden Zeitschrittes von Molekulardynamiksimulationen ein. Zuvor wird das System mit einer Energieminimierung (Kapitel 2.5) auf den Boden des nächstgelegenen lokalen Minimums getrieben. Die Analyse geht dabei von einem harmonischen Potentialtopf aus [95, 96]. Die NMA ist der Flaschenhals des MMPBSA Verfahrens, denn sie ist sehr zeitaufwendig und beansprucht viel Platz im Speicher des Rechners. Aus diesem Grunde wird die NMA bei vielen Berechnungen oft übergangen [97]. Die Rechtfertigung dazu erfolgt in weiteren Arbeiten die zeigen dass die Bestimmung der Normalmoden nicht immer notwendig ist [98, 99, 100, 101, 102].

2.6.3 Fittingmethoden

In der biomolekularen Modellierung beschreibt das Fitting ein Verfahren, bei dem zwei oder mehrere Konformationen zueinander ausgerichtet werden. Ein solcher Algorithmus erfordert ein numerisches Maß für die Unterschiede zwischen den Konformationen und überlagert am Ende die Atome oder Atomgruppen optimal. Das wohl wichtigste Maß in diesem Zusammenhang ist die mittlere quadratische Abweichung, auch RMSD (root mean square deviation) genannt. Es sind viele Verfahren entwickelt worden, um diese zu errechnen. Das Ziel der RMSD Berechnungen ist es, eine Ausrichtung mit einem relativen Minimum zu finden. Ferro und Hermans verwenden ein iteratives Verfahren, bei dem ein Molekül relativ zum anderen bewegt wird und so schrittweise die RMSD reduziert wird [103]. Bei andere Algorithmen, wie bei dem von Kabsch [104], wird die beste Lösung über die folgende Gleichung ermittelt:

$$RMSD_j = \sqrt{\frac{1}{N} \sum_i^N (x_{j,i} - x_{R,i})^2} \quad (2.24)$$

Hier ist N die Anzahl der Atome mit denen die RMSD bestimmt werden soll. Der Abstand zwischen den Koordinaten, von Atom i , in einer entsprechenden Struktur j und einer Referenzstruktur R ist $x_{j,i} - x_{R,i}$. Ähnliche Berechnungen wie RMSD Berechnungen sind RMSF (root mean square fluctuation) Berechnungen. Jedoch wird beim RMSD für jeden Zeitschritt in einer Trajektorie ein Wert ermittelt, beim RMSF existiert nur ein Wert für alle Zeitschritte in einer Trajektorie.

$$RMSF_i = \sqrt{\frac{1}{T} \sum_j^T (x_i(t_j) - \bar{x}_i)^2} \quad (2.25)$$

Gleichung 2.25 beschreibt den RMSF. T ist dabei die Anzahl der betrachteten Strukturen, x_i die Atomkoordinaten und \bar{x}_i die Atomkoordinaten über ihrem zeitlichen Mittel [105]. Die RMSF eines Proteins kann auch mit dem experimentellen B-Faktor aus der Röntgenstrukturanalyse verglichen werden. Aufgrund der unterschiedlichen Definitionen von RMSF und B-Faktor, sollte eine Eins-zu-Eins-Entsprechung nicht erwartet werden. Allerdings sollten die Ergebnisse der Simu-

lation mit denen der Kristallographie korrelieren.

2.6.4 Clusteranalyse

Programme der biomolekularen Modellierung erzeugen eine große Menge an Daten, die verarbeitet und analysiert werden müssen. Algorithmen für Molekulardynamiksimulationen und für Docking, erzeugen Konformationen, die identisch oder ähnlich sind. Deshalb ist es oft sinnvoll, die Daten auf einen repräsentativen Datensatz zu reduzieren. Dieser kann anschließend in Analysen ausgewertet werden. Eine Möglichkeit zur Datenreduktion ist die Clusteranalyse, in der Gruppen von vergleichbaren Objekten zusammengefasst werden und aus denen dann ein repräsentativer Vertreter ausgewählt wird.

Zur Verifizierung einer Clusteranalyse wird ein Ähnlichkeitsmaßstab für die Datensätze benötigt. Es kann keine eindeutig richtige Methode benannt werden, da die Auswahl der Clusteranalyse unter anderem abhängig vom Auswertalgorithmus für den entsprechenden Datensatz ist [106].

Bei dem Vergleich von Konformationen ist die RMSD ein sinnvoller Maßstab. Alternativ können auch die Winkel zwischen den Konformationen bestimmt werden. Außerdem gibt es weitere Möglichkeiten, um Distanzen zu beschreiben. Die resultierenden Cluster der obigen Beispiele (RMSD und Torsionswinkel) können sehr stark variieren. Dies ist auf den Leverage Effekt bei der Verwendung von Torsionswinkeln zurückzuführen. Er tritt bei kleinen Änderungen der Torsionswinkel in der Mitte des Moleküls auf und kann zu großen Änderungen am Rande des Moleküls führen. RMSD Daten produzieren im Gegensatz zu Datensätzen auf der Grundlage von Torsionswinkeln Cluster mit Formen [6].

Eine Gruppe von Clusteralgorithmen sind die *linkage* Algorithmen [107]. Diese Algorithmen berechnen zuerst den Abstand zwischen jedem Paar von Konformationen. Somit enthält der Datensatz zu Beginn die gleiche Menge an Clustern und Konformationen. Danach werden die Cluster in jedem Schritt um eins reduziert, da die zwei Cluster mit stärkster Ähnlichkeit zu einem Cluster zusammengefasst werden. Das Clustering ist beendet, wenn alle Konformationen in einem Cluster zusammengefasst sind, wenn ein definiertes Clusterminimum erreicht wird, oder wenn ein definierter Abstand zwischen dem nächsten Paar von Clustern über-

schritten wird. Bei diesen Algorithmen handelt es sich um *agglomerative* Methoden. Im Gegensatz dazu starten die *distributiven* Methoden aus einem einzigen Cluster heraus [106]. Die Konformation, die am wenigsten von der Durchschnittsstruktur eines Clusters abweicht, wird als repräsentative Struktur durch den Algorithmus ausgegeben.

Der Abstand zwischen zwei Clustern kann auf mehrere Arten berechnet werden. Beispiele sind die *single linkage* Methode, die *complete linkage* Methode, die *average linkage* Methode. Bei der *single linkage* Methode wird der kürzeste Abstand durch den Vergleich aller Distanzen zwischen allen Mitgliedern der beiden Cluster bestimmt [108]. Die *complete linkage* Methode ist das Gegenteil und sucht das am weitesten entfernte Paar von Objekten. Die *average linkage* Methode errechnet den Durchschnitt der Ähnlichkeiten. Im Hinblick auf ihre ähnliche Arbeitsweise werden diese drei Methoden *hierarchische* Clusterverfahren genannt. Sie haben den Vorteil, einfach programmierbar zu sein und erzeugen, beziehend auf die Reihenfolge, in der die Objekte gespeichert werden, unabhängige Cluster. Für die Berechnung wird jedoch eine Äquivalenzrelation auf der Klasse der quadratischen Matrizen aufgestellt, die ihre Anwendbarkeit bei großen Datenmengen begrenzt. Die *single linkage* Methode neigt zur Erzeugung eines einzelnen ausgedehnten Clusters. Dies macht wiederum die Anwendung einer alternativen Clustermethode erforderlich [106]. Zu einem hierarchischem Clustering kann ein Dendrogramm dargestellt werden. Es zeigt anschaulich die Beziehung zwischen den Positionen der Daten und zu jedem Zeitpunkt den Wert und die Mitglieder eines jeden Clusters. Auf diese Weise kann die darunterliegende Struktur als Information für die optimale Anzahl von Clustern verwendet werden. Mit einer Linie über das Dendrogramm kann die Clustergröße auf einen bestimmten Grenzwert gehalten werden. Die Entscheidung über die Clustergröße ist subjektiv, da ein kleiner Grenzwert zu einer großen Zahl von Clustern mit nur einem Mitglied führt, während ein großer Grenzwert zu ausgedehnten Clustern führt [6].

Der Jarvis-Patrick Algorithmus ist ein Beispiel für ein nicht-hierarchisches Clusterverfahren [109]. Er ist als *nearest neighbors* Ansatz bekannt. Zwei Konformationen werden hier demselben Cluster zugeordnet, wenn sie in der Liste der m nächsten Nachbarn des jeweils anderen sind oder wenn sie p (wobei $p < m$) nächste Nachbarn gemeinsam haben. Konformationen können in Clustern plat-

ziert werden und Cluster können ohne hierarchische Beziehungen verschmolzen werden. Das Verfahren kann auch durch die Positionen der einzelnen Konformation innerhalb der Nachbarliste erweitert werden. Ein Vorteil des Algorithmus ist es, sehr große Datenmengen Clustern zu können. Eine häufige Anwendung ist die Auswahl von repräsentativen Molekülen aus großen chemischen Datenbanken [110].

2.6.5 Signifikanzanalyse

Die Analyse von Molekulardynamiksimulationen, speziell zur Strukturverfeinerung, ist derzeit nicht standardisiert. Noch vor einigen Jahren wurden Simulationen manuell, mit Tools wie PyMol oder VMD analysiert. Mit der Entwicklung zunehmender Rechnerleistungen wurden effektivere Ansätze möglich. Simulationseigenschaften, wie RMSD (Kapitel 2.6.3), konnten aufgezeichnet und intuitiv analysiert werden [111]. Durch die Entwicklung umfangreicherer Simulationstechniken sind auch diese Methoden mittlerweile nicht mehr ausreichend. Neue Ansätze finden repräsentative Strukturen nur mit moderatem Erfolg. Zur Zeit ist es allgemein üblich die letzte Struktur einer Simulation als die repräsentative Struktur anzunehmen [112, 113, 114]. Es muss jedoch beachtet werden, dass Molekulardynamiksimulationen nie ein vollständiges Gleichgewicht erreichen [115, 116, 117, 113]. Darüber hinaus steht die Vorgehensweise in Konflikt mit der Ergodenhypothese, die auf einer repräsentativen Probenahme basiert. Sich daraus ergebende Probleme können jedoch durch statistische Methoden umgangen werden.

Die SAM (significance analysis of microarrays) Analyse [118], ist ein modifizierter T-Test, um eine Alphafehler-Kumulierung zu vermeiden. Robert Tibshirani entwickelte SAM zu Beginn der Gen-Expression-Microarray-Experimente. Hier verglich er Daten von nicht geheilten und geheilten Hautkrebspatienten in Abhängigkeit von der angewendeten Therapie [119]. Später erkannte er, dass dieser Test generelle Anwendung finden kann und entwickelte mit Balasubramanian Narasimhan ein Rechnerprogramm zur allgemeinen anwendbaren Analyse der Daten [120]. Holger Schwender entwickelte später das *siggenes* [121] Paket für die R Distribution, welches auch in der vorliegenden Arbeit verwendet wird. Nach-

folgend ist eine qualitative Beschreibung dieser Methode zu finden. Für eine exakte Beschreibung und Herleitung der Zusammenhänge wird das Buch von Hastie, Tibshirani und Friedman, *The Elements of Statistical Learning* [122, Kapitel 18.7], empfohlen. Wie die Methode auf Molekulardynamiksimulationen übertragbar ist, wird in Kapitel 3.3.2 gezeigt.

Die Eingangsdaten haben die Form einer $m \times n$ Matrix, wobei m als Gene und n als Beobachtungen mit einer entsprechenden Antwort gesehen werden. Um die Nullverteilung abzuschätzen wird eine Anzahl von Permutationen voreingestellt.

1. Für jedes Gen, wird ein d-Wert berechnet (analog zu einer T-Statistik). Dies ist der *beobachtete d-Wert* für dieses Gen.
2. Die Gene werden in aufsteigender Reihenfolge ihrer d-Werten sortiert.
3. Ein Zufallsprinzip mischt die Werte der Gene, zwischen den Gruppen A und B. Nach der Umgruppierung haben beide Gruppen die ursprünglich vorhandene Anzahl von Elemente. Für jedes randomisierte (zufällig ausgewählte) Gen wird ein d-Wert berechnet.
4. Die permutierten d-Werte der Gene werden in aufsteigender Reihenfolge sortiert.
5. Die Schritte 3 und 4 werden mehrmals wiederholt, so dass jedes Gen eine bestimmte Anzahl von randomisierten d-Werten besitzt. Der Durchschnitt der randomisierten d-Werte ist der *erwartete d-Wert* für jedes Gen.
6. Die *beobachteten d-Werte* werden gegen die *erwarteten d-Werten* aufgetragen (Abbildung 2.1).

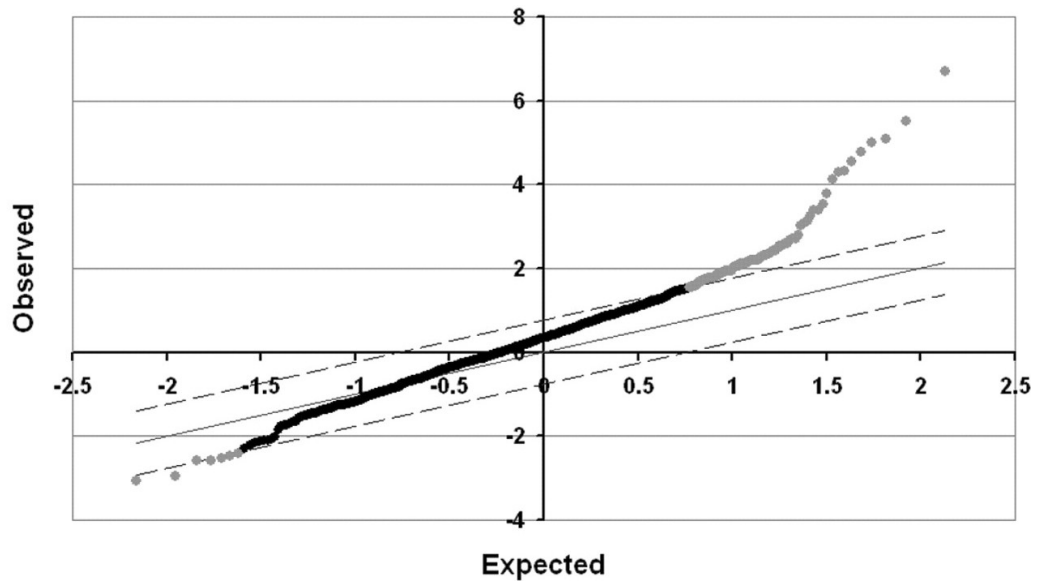


Abbildung 2.1: Exemplarischer SAM Plot [123]. Auftragung von beobachteten d-Werten (hier **Observed**) gegen erwarteten d-Werten (hier **Expected**). Die gestrichelten Linien zeigen die Schwellenwerte Δ . Die Werte oberhalb bzw. unterhalb von Δ sind grau gefärbt und gehören zu den unterschiedlich exprimierten Genen S_{Δ} .

Für einen Satz von diskreten Schwellenwerten Δ wird eine obere und eine untere Grenze bestimmt. Diese Grenzwerte definieren eine Reihe von Gene S_{Δ} die unterschiedlich exprimiert wurden. Als Fehlerabschätzung wird eine *False Discovery Rate* für jedes S_{Δ} bestimmt.

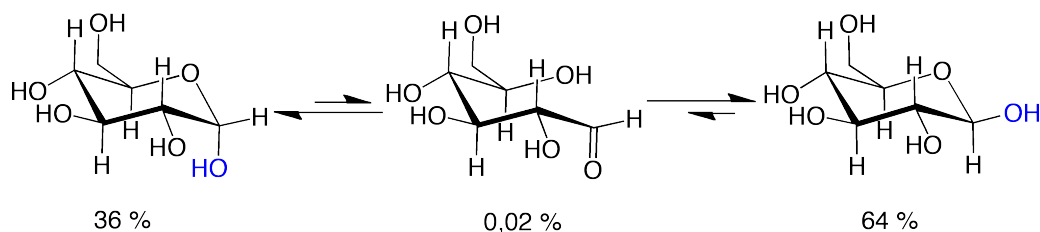


Abbildung 3.1: Die Epimere von D-Glucose [125]. Die Kohlenstoffatome der D-Glucosepyranose werden im Uhrzeigersinn gezählt beginnend mit dem C-Atom der blau dargestellten Hydroxygruppe. Die α -D-Glucopyranose (links) trägt dabei seine blau markierte Hydroxygruppe in der axialen Position, während die β -D-Glucopyranose (rechts) seine blau markierte Hydroxygruppe in einer äquatorialen Position trägt. In beiden Fällen ist die Hydroxygruppe an C1 des Pyranosesystems koordiniert.

3 Systeme und Anwendungen

3.1 Modellierung spezifischer GDH Isoenzyme

3.1.1 Einführung

Das Genom des extremen thermoacidophilen Archaeon *Sulfolobus solfataricus* umfasst verschiedenen Kandidaten des GDH Enzyms. Der Kandidat SSO_3003 ist dabei als Kristallstruktur verfügbar [22, PDB ID: 2cdc, 2cdb, 2cda, 2cd9]. Dieser Kandidat hat eine breite Monosaccharidspezifität, mit einer Vorliebe für äquatoriale Hydroxygruppen, insbesondere an den Zuckerpositionen C2 - C4 [124].

Abbildung 3.1 zeigt die zu 99 % in Lösung vorkommende D-Glucopyranose, mit ihrem Epimerenverhältnis von $\alpha : \beta = 36 : 64$. Blau eingezeichnet ist dabei die C1-Hydroxygruppe. Im α -Epimer ist sie dabei in axialer Position angeordnet, im β -Epimer hingegen in äquatorialer Position. Die Kohlenstoffatome in der D-Glucopyranose werden von hieraus im Uhrzeigersinn weiter gezählt. C6 liegt dabei außerhalb des Ringes an C5 gebunden, welcher Teil des Ringes ist. In der β -D-Glucose sind alle Hydroxygruppen äquatorial angeordnet. Xylose ist ein Zucker mit der gleichen Ringkonformation wie Glucose, aber ohne $-\text{CH}_2-\text{OH}$ Gruppe an C5, welche durch Wasserstoff ersetzt ist. In dieser Arbeit wird SSO_3003 als Referenzstruktur verwendet. Tabelle 3.3 zeigt die von Milburn [22] gefunde-

3.1 Modellierung spezifischer GDH Isoenzyme

nen Interaktionen, von Glucose und Xylose mit SSO_3003. Sie spiegelt die schon angesprochene Bevorzugung der äquatorialen Hydroxygruppen an C2 - C4 [124] wieder.

Im Gegensatz zu SSO_3003, ist der Kandidat SSO_3024 bei jedem Monosaccharid inaktiv, wobei SSO_3042 fast ausschließlich die Konvertierung von Glucose katalysiert. Dabei wurde der Kandidat SSO_3024 von Dr. Eleni Theodosiou und der Kandidat SSO_3204 von Dr. Patrick Haferkamp analysiert und exprimiert. Um die experimentellen Ergebnisse dieser Kandidaten zu rationalisieren, werden Strukturmodelle von SSO_3024 und SSO_3204 erstellt. Zusätzlich werden elektrostatische Oberflächen von SSO_3003 und SSO_3024 hergestellt und miteinander verglichen. Interessanterweise reflektieren die elektrostatischen Oberflächen die Sequenzunterschiede. Darüber hinaus wird ein molekulares Docking mit einem Modell von SSO_3024 und verschiedenen Monosacchariden durchgeführt.

3.1.2 Material und Methoden

Generieren der Strukturmodelle Die dreidimensionale Struktur von SSO_3204 wird unter Verwendung des folgenden Verfahrens modelliert. Um mögliche Strukturvorlagen zu finden, werden zunächst die Aminosäuresequenzen bei mGenTHREADER [126] hochgeladen. Die jeweils beste Vorlage, die auf diese Weise identifiziert wird, ist die Kristallstruktur der Glucose Dehydrogenase SSO_3003 [22, PDB ID: 2cdb, 2cdc]. Mit den Aminosäuresequenzen von 28 Glucose Dehydrogenasen, einschließlich dem entsprechendem Isoenzym und SSO_3003, wird jeweils ein Sequenzalignment [127] durchgeführt. Danach werden die Strukturen von SSO_3204 und SSO_3024 auf Grundlage ihres jeweiligen Sequenzalignments und der Struktur von SSO_3003 modelliert [128] und anschließend mit PyMOL [129] visualisiert. Um die Robustheit der Modelle zu testen wird das Verfahren mit der Glucose Dehydrogenase des extremen halophilen *Haloferax mediterranei* [130] als Vorlage wiederholt. Die Struktur dieser Glucose Dehydrogenase wird als zweitbestes Ergebnis in der mGenTHREADER Suche gefunden. Abgesehen von kleinen Verschiebungen in den Atompositionen sind diese Modelle identisch mit den ersten Modellen.

3.1 Modellierung spezifischer GDH Isoenzyme

Elektrostatistische Oberfläche Die elektrostatischen Oberflächen der Strukturen von ESO_3003 und ESO_3024 werden mit PyMol [129] und dem APBS [131] Plug-in berechnet und visualisiert. Die Strukturen werden mit pdb2pqr [132, 133], des Implementierungen der Amber99 Kraftfeldes und propKa [134] vorbereitet. PropKa wird mit einem pH-Wert von 7,5 verwendet. EpiTopsy wird für eine relative Quantifizierung der elektrostatischen Oberflächen [135] verwendet.

Docking von Monosacchariden Die gedockten Monosaccharide waren β -D-Xylopyranose, α -D-Xylopyranose, Glucose, 6-Desoxy-Glucose und 2-Desoxy-Glucose. Die ersten drei Liganden werden aus den Kristallstrukturen von ESO_3003 [22, PDB ID: 2cdb, 2cdc] extrahiert. Hier liegt die Glucose mit zwei Orientierungen für den Sauerstoff an C6 vor. Die Monosaccharide werden mit der AutoDock [136] Implementierung von YASARA [137] gedockt. Dabei werden 999 Durchläufe vorgenommen und die Ligandenpositionen werden geclustert, wenn der RMSD des Liganden kleiner als $0,5 \text{ \AA}$ ist. Die Cluster werden danach mit dem Amber03 Kraftfeld in YASARA nach guten und schlechten Positionen sortiert. Der Mittelwert der zehn besten Clusterergebnisse, für jeden Modell-Monosaccharid-Komplex, werden berechnet und den anderen Modell-Monosaccharid-Komplex gegenübergestellt.

3.1.3 Ergebnisse und Diskussion des Isoenzyms SSO_3204

Ein Vergleich der Bindetaschen der Kristallstruktur von SSO_3003 und der berechneten Modelle von SSO_3204 zeigt eine Reihe von Unterschieden zwischen den Enzymen. Dies schließt Aminosäuren ein, welche von Milburn et al. [22] als kritisch für die Interaktion mit den Monosacchariden identifiziert wurden (siehe auch Tabelle 3.1 und Abbildung 3.2).

Im Folgenden werden diese Unterschiede und ihre voraussichtlichen Auswirkungen auf die Bindung von Glucose und Xylose diskutiert. Der erste wichtige Unterschied zeigt sich von N89 in SSO_3003 zu V93 in SSO_3204. Dies führt wahrscheinlich zu einem Verlust der Wasserstoffbrückenbindung zwischen dem Pyranosering des Monosaccharides und dem Enzym, begründet in einer schwächeren Affinität von Glucose und Xylose resultiert. Zweitens existieren Mutationen von

3.1 Modellierung spezifischer GDH Isoenzyme

Q150 nach E154 und von D154 nach N158. Diese Mutationen könnten als neutral in Bezug zu der Substratbindung betrachtet werden, da die funktionellen Gruppen in Summe identisch bleiben und die Gesamtladung nicht verändert wird. Eine drastische Veränderung ist hingegen die Mutation des neutralen bzw. basischen H297 in SSO_3003 auf das saure E294 in SSO_3204. Diese Interaktion wird von einer benachbarten Polypeptidkette des tetrameren Enzymes erzeugt. Da das Histidin $N\epsilon 2$ Atom an einer Wasserstoffbrückenbindung zur C6-Hydroxygruppe der Glucose beteiligt ist, ist es naheliegend, dass die längere und saure Glutamatgruppe zu einer stärkeren Wasserstoffbrücke mit der gleichen Hydroxygruppe führt. Dies könnte die schwächere Wechselwirkung mit dem Pyranosering der Glucose durch die oben diskutierte Mutation von N89 nach V93 kompensieren, so dass das Isoenzym SSO_3204 trotzdem noch Glucose binden kann. Im Gegensatz dazu hat

Tabelle 3.1: Aminosäuren der Bindetaschen von SSO_3003 und SSO_3204.

Vergleich der Aminosäuren in den Bindetaschen der Kristallstruktur SSO_3003 [22, PDB ID: 2cdb] und dem Modell SSO_3204. Aufgeführt sind dabei alle Aminosäuren in einem Umkreis von 5 Å eines gebundenen Glucosemoleküls. Die Sternchen markieren dabei die nach Milburn et al. wichtigsten Wechselwirkungen [22]. Die Aminosäuren in der letzten Reihe werden durch eine benachbarte Polypeptidkette des tetrameren Enzyms bereitgestellt.

SSO_3003	SSO_3204
C39	C40
R41	Y42
H66	H67
N89*	V93
R90	R94
E114*	E118
I117	I121
Q150*	E154
D154*	N158
F279	–
V306	V303
N307*	N304
G308	A305
NAP	NAP
H297*	E294

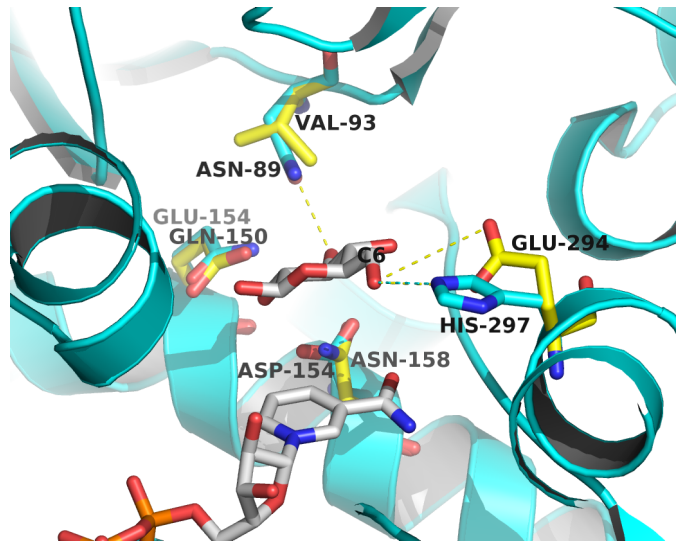


Abbildung 3.2: Überlagerung der Bindetaschen von SSO_3003 und SSO_3204. Die Überlagerung der Bindetaschen mit Glucose (Mitte, grau und rot), von der Kristallstruktur SSO_3003 (cyan) und dem Modell SSO_3204 (gelb).

Xylose keine C6-Hydroxygruppe, und kann somit nicht den Verlust der Wasserstoffbrückenbindung an N89 kompensieren. Die Affinität von Xylose an das Enzym sollte daher wesentlich von SSO_3003 nach SSO_3204 reduziert sein. Dies ist konsistent mit den experimentellen Beobachtungen, da beide Enzyme Glucose als Substrat verwenden, während nur SSO_3003 Xylose verwendet.

3.1.4 Ergebnisse und Diskussion des Isoenzyms SSO_3024

Dieses Modell, wie auch das vorige zeigen beide die charakteristischen Aminosäuren für die Coenzymbindung [22] der entsprechenden Superfamilie [15]. Im Gegensatz dazu weisen beide Modelle Unterschiede zur Referenz [22, SSO_3003] in der Monosaccharidbindungsstelle auf.

Es ist zu beachten, dass die folgenden Ergebnisse die Quartärstruktur des Enzyms vernachlässigen. Dies könnte ein Problem sein, da Dr. Eleni Theodosiou in persönlichen Gesprächen eine α_3 Quartärstruktur vermutete. Die Hauptvorlage SSO_3003 zeigte hingegen ein Tetramer als Quartärstruktur. Eine Vorlage mit einem geeigneten Trimer war nicht verfügbar.

3.1 Modellierung spezifischer GDH Isoenzyme

Unterschiede in den Bindungstaschen Tabelle 3.2 stellt die Unterschiede in der Bindungstasche des Modells SSO_3024 und der Referenzstruktur SSO_3003 in einem Umkreis von 5 Å um ein gebundenes Substrat dar. Der erste Unterschied ist in der hydrophoben Aminosäure L50 zu finden, welche durch die basische Aminosäure R70 ersetzt wird. In diesem Modell blockiert dabei die Seitenkette von R70 teilweise den Eingang der Bindungstasche. Darüber hinaus ist es möglich, dass R70 durch T137 und T61, die etwa 3 Å entfernt sind, in seiner Position fixiert wird (Abbildung 3.3). Damit würde R70 als Pfortner der Bindungstasche agieren, eine Funktion, die es wahrscheinlich mit L50 in der Referenzstruktur nicht gab.

Der zweite Unterschied beschreibt die zwei basischen Aminosäuren K110 und R90. Diese Mutation scheint unbedeutend, weil sie die chemischen Eigenschaften nicht beeinflusst. Jedoch könnten anderer Aminosäuren (E114, N307) durch diese Mutation in ihren Positionen destabilisiert sein [22]. Dies würde es schwierig machen, Monosaccharide für einen hohen enzymatischen Umsatz in der Bindetasche zu fixieren.

Die dritte Mutation beschreibt einen Wechsel von E114 in SSO_3003 zu V134 in SSO_3024. Dies führt zu einem Verlust von zwei in Tabelle 3.3 beschriebenen Wasserstoffbrückenbindungen. Es ist unwahrscheinlich, dass diese durch D326, fünfter Eintrag in der Tabelle 3.2 kompensiert werden können, da die Längen der Seitenketten unterschiedlich sind. Des Weiteren sind D326 und E114 nur in der gleichen Region positioniert, nicht aber an identischen Positionen der

Tabelle 3.2: Aminosäuren der Bindetaschen von SSO_3024 und SSO_3003. Unterschiede in den Bindungstaschen des Modells SSO_3042 und der Referenzstruktur SSO_3003, in einem 5 Å Radius eines Substrates.

SSO_3024	SSO_3003	
Aminosäure	Aminosäure	Funktion[22]
R70	L50	—
K110	R90	stabilisiert E114 and N307
V134	E114	C4-/C6-hydroxyl
T137	I117	—
0 D326	V306	—
I317	H297	C6-hydroxyl (Alternative Position)

3.1 Modellierung spezifischer GDH Isoenzyme

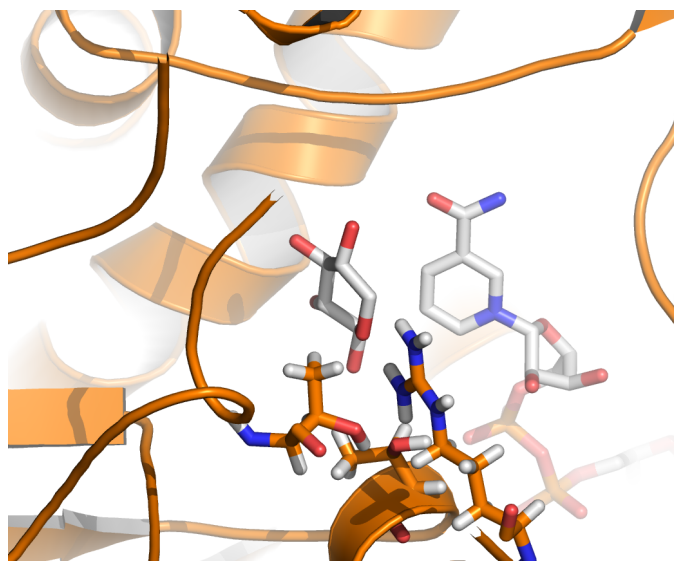


Abbildung 3.3: R70 als Pförtner in SSO_3024. Konfiguration von R70, T61 und T137 im Modell von SSO_3024. Die Struktur ist in orangefarbenem Cartoon mit R70, T61 und T137 als Stäbchenmodell angezeigt. Das Coenzym und das Monosaccharid werden als weiße Stäbchenmodelle gezeigt.

Tabelle 3.3: Wichtigste GDH Wechselwirkungen. Nach Milburn et al. [22], wichtigste atomare Wechselwirkungen zwischen Glucose bzw. Xylose und SSO_3003.

Monosaccharidposition	GDH (SSO_3003)
C1	Nicotinamid C4
O2	D154 O δ 2 Q150 N ϵ Q150 O ϵ
O3	D154 O δ 1 D154 O δ 2 N307 N δ Q150 N ϵ N89 N δ
O4	E114 O ϵ 2 N307 N δ
O6	E114 O ϵ 1
O6 (Alternative Position)	H297 N ϵ 2

3.1 Modellierung spezifischer GDH Isoenzyme

Bindungstaschen. Da sich hierdurch sowohl der Winkel als auch die Länge der Wasserstoffbrückenbindung verändert, wäre eine Präferenz für äquatoriale C4-Hydroxygruppen in SSO_3024 nicht mehr gegeben.

Tabelle 3.3 zeigt, dass C6-Monosaccharide zwei Möglichkeiten zur Bindung haben. Zum einen über eine Wasserstoffbrückenbindung zu E114, oder zum anderen über eine alternative Wasserstoffbrückenbindung zu H297. Die letzte Zeile von Tabelle 3.2 zeigt dass diese Möglichkeit nicht besteht. Wenn überhaupt, würde I317 an dieser Position sitzen, was aber durch die vermutete Quartärstruktur unwahrscheinlich ist.

Zusammenfassend kann festgehalten werden, dass es möglich ist, dass ESO_3024 keine Substrate mit C6-Hydroxygruppen stabilisiert. Darüber hinaus ist die Fähigkeit, C4-Hydroxygruppen zu stabilisieren um etwa die Hälfte reduziert. Es ist daher zu erwarten, dass die Möglichkeit einer Substratbindung ohne C4- bzw. C6-Hydroxywechselwirkung und dem Auftreten eines Pfortners sehr gering ausfallen wird.

Unterschiede im elektrostatischen Potential Neben der Beeinflussung der direkten Protein-Substrat-Wechselwirkungen haben die oben beschriebenen Mutationen auch Einfluss auf das elektrostatische Potential in und um die Bindungstasche. Abbildung 3.4 zeigt dazu die elektrostatische Oberfläche von SSO_3003 und SSO_3024 bei $\pm 2kT/e$, wobei k die Boltzmann-Konstante, T die Temperatur in Kelvin und e die Elementarladung ist. Ein offensichtlicher Unterschied zwischen den beiden Oberflächen ist die große rote Blase in der Referenzstruktur (obere Struktur in Abbildung 3.4) über der Bindungstasche. Monosaccharide werden förmlich in die Bindungstasche gezogen, wenn sie diese Blase berühren. Im Gegensatz dazu werden Monosaccharide die Bindungstasche des Modells (untere Struktur in Abbildung 3.4) nicht finden. Dies wird durch Experimente bestätigt. Um beide Oberflächen sowie den Einfluss der Mutationen auf diese Oberflächen besser miteinander vergleichen zu können, wird eine relative Quantifizierung der elektrostatischen Bindungsoberflächen in einem Umkreis von 10\AA von einem Substrat berechnet. Sie basiert auf dem Mittelwert aus 10 konformativ verschiedenen SSO_3003 Strukturen. Dieser Mittelwert wird als Referenz und somit als 1 in Abbildung 3.5 definiert.

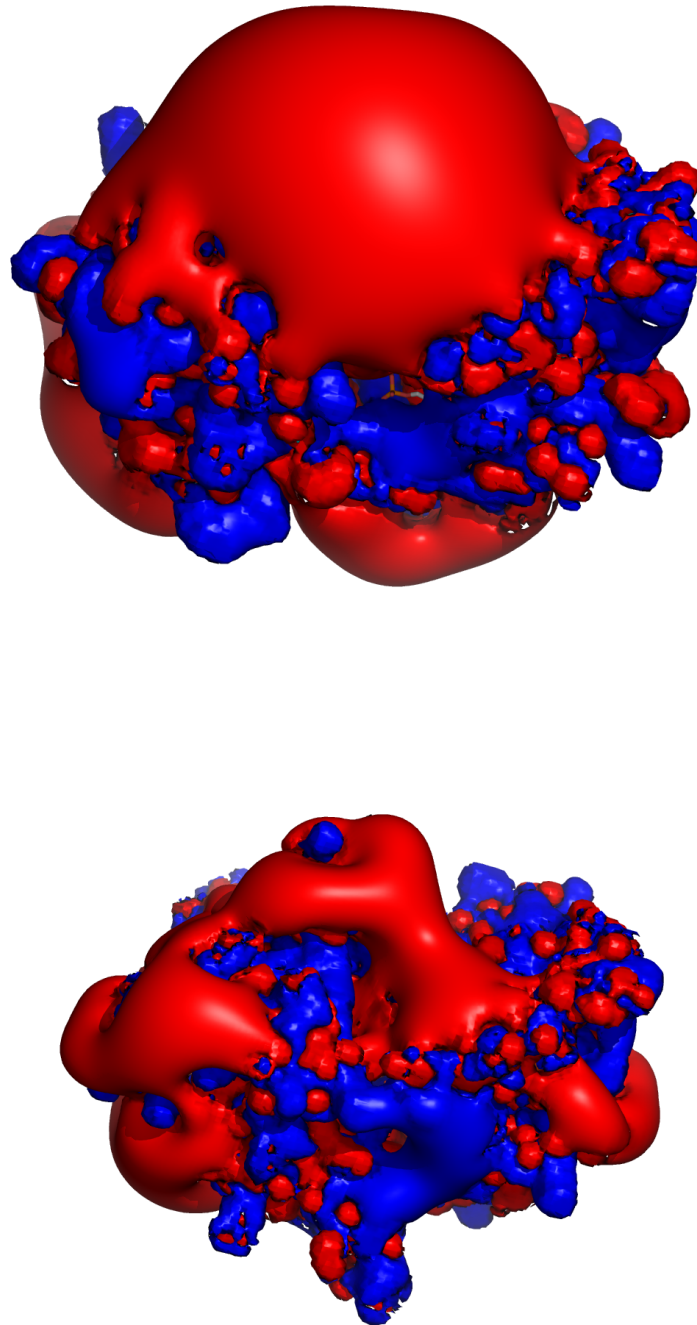


Abbildung 3.4: Elektrostatisches Potential von SSO_3003 und ESO_3024. Elektrostatisches Potential der Referenzstruktur SSO_3003 (oben) und dem Modell SSO_3024 (unten) bei $\pm 2kT/e$. In der Referenzstruktur ist unterhalb der großen roten Blase die vom Substrat weiter entfernte Nucleotid-Phosphatgruppe des Co-Enzyms zu erkennen.

3.1 Modellierung spezifischer GDH Isoenzyme

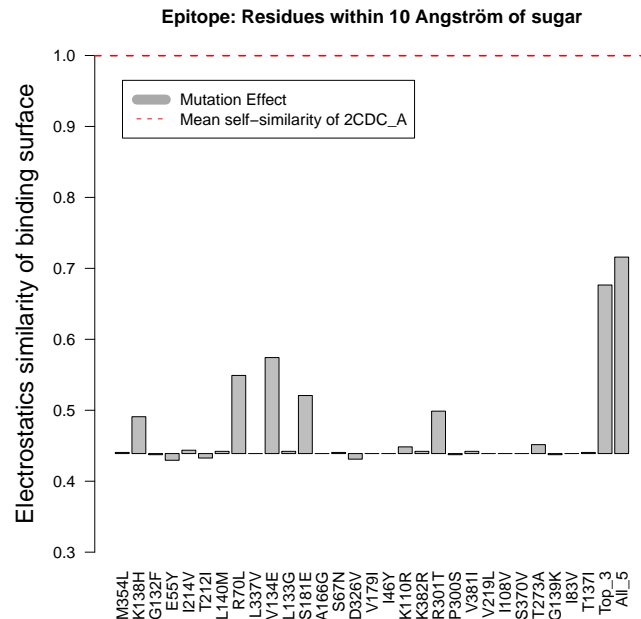


Abbildung 3.5: Elektrostatische Ähnlichkeit zwischen der Bindungsoberfläche von ESO_3003 und ESO_3024. Die elektrostatische Ähnlichkeit wurde hier in einem Radius von 10 Å um ein koordiniertes Substrat berechnet. Auf der äquatorialen Achse sind alle Mutationen in einem Radius von 15 Å um ein koordiniertes Substrat aufgeführt. Der Einfluss der Mutationen auf die elektrostatische Bindungsoberfläche ist grau dargestellt. Ausgangspunkt für diese Balken ist eine Ähnlichkeit von etwa 44 % des Modells, ESO_3024, mit der Referenz, wobei ESO_3003 als Referenz auf 100 %, bzw. 1, gesetzt worden ist.

Für das Modell von ESO_3024 wird ebenfalls ein Mittelwert aus 10 Konformeren generiert und mit der Referenz verglichen. Es zeigt sich dabei eine Ähnlichkeit der elektrostatischen Oberflächen von 44 %. Alle unterschiedlichen Aminosäuren in einem Radius von 15 Å um ein Substrat werden daraufhin zur Referenz mutiert, es werden Konformere gebildet und wiederum mit der Referenz verglichen. Beispielsweise wird die oben diskutierte Mutation von K110 in SSO_3024 nach Leucin durchgeführt und als K110L in Abbildung 3.5 dargestellt. Es zeigt sich, dass diese Mutation nur geringen Einfluss auf die elektrostatische Ähnlichkeit der Oberflächen hat. Andere Mutationen, wie K138H, R70L, V134E, S181E und R301T zeigen hingegen einen stärkeren Einfluss.

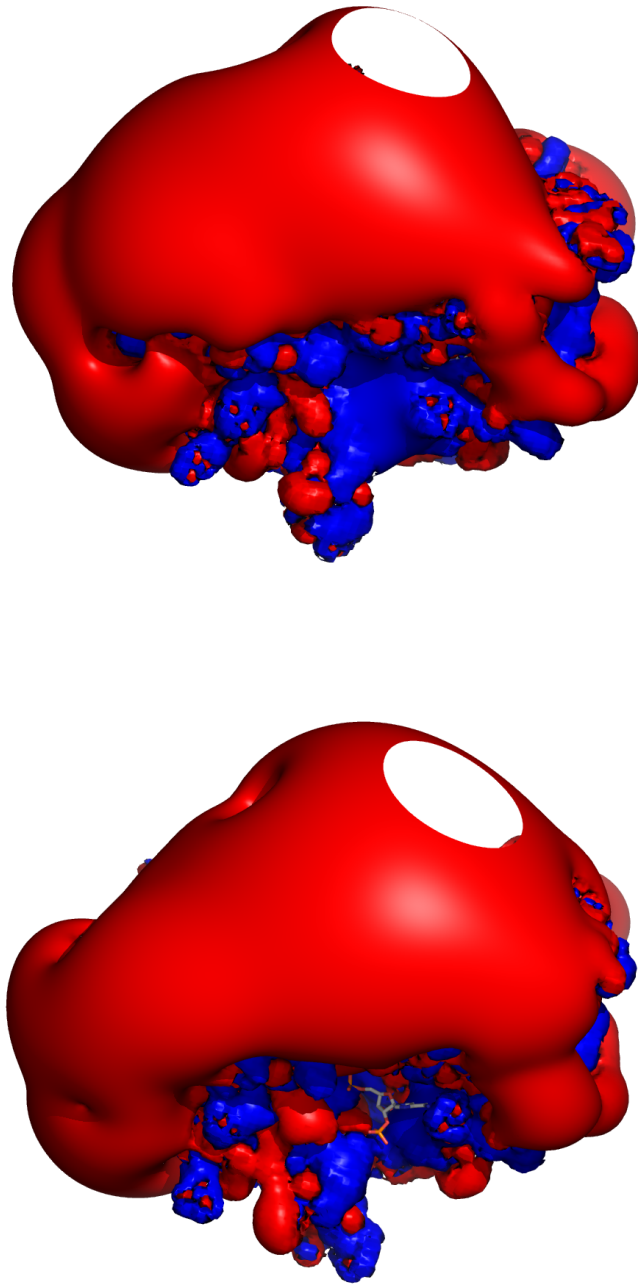


Abbildung 3.6: Elektrostatisches Potential der TOP_3 und All_5 Mutationen. Die Abbildung zeigt die elektrostatische Oberfläche der TOP_3 (oben) und All_5 (unten) Mutationen bei $\pm 2kT/e$. Diese haben den größten Einfluss auf die elektrostatische Ähnlichkeit der Oberflächen von ESO_3003 und ESO_3024 (Abbildung 3.5).

3.1 Modellierung spezifischer GDH Isoenzyme

Alle diese Mutationen haben einen Einfluss von 5 – 15 % auf die elektrostatischen Oberflächen. Eine Mutation aller fünf Aminosäuren führt sogar zu einer Ähnlichkeit von rund 75 % und ist in Abbildung 3.5 als All_5 zu finden, TOP_3 bezieht sich hingegen auf die Mutationen R70L, V134E und S181E. Interessanterweise findet sich die oben diskutierte R70L Pförtnermutation auch unter diesen TOP_3 bzw. All_5. Abbildung 3.6 zeigt die Oberflächen der TOP_3 und All_5 Mutationen. In dieser Abbildung ist zu erkennen, dass die rote Blase, unter der die Bindungstasche liegt, von ESO_3003 (obere Abbildung 3.4) nachgebildet werden kann. Auch wenn sie durch die TOP_3 bzw. All_5 Mutationen überproportional erscheint, kann der Effekt, den einzelne Punktmutationen auf die elektrostatische Oberfläche von Proteinen haben, klar gezeigt werden. Wie in Abbildung 3.5 durch die E55Y Mutation angedeutet ist, können auch Mutationen eingeführt werden, die eine Ausbildung der roten Blase regulieren.

Docking der Monosaccharide Das Docking von Liganden in modellierte Strukturen ist eine nützliche Methode, sowohl zu Verifizierung des Modells als auch zur Generierung neuer Erkenntnisse, wie z.B. dem Finden neuer Liganden. Zunächst muss jedoch getestet werden, ob das Docking für dieses System eine brauchbare Methode ist. Hierzu werden bekannte Umsetzungsgeschwindigkeiten des Enzyms reproduziert. Gelingt dies, können unbekannte Modellstrukturen untersucht werden. Diese beiden Schritte werden deshalb im Folgenden untersucht.

Tabelle 3.4 zeigt die Ergebnisse für ESO_3003 und ESO_3024, in beiden Fällen mit den TOP_3 Mutationen. Aus ihr geht eine gute Übereinstimmung der relativen Umsetzungsgeschwindigkeiten der Enzyme (nach Lamble et al. [15]) und den Dockingenergien hervor. Darüber hinaus zeigen die Dockingergebnisse keinen großen Einfluss des Kristallwassers aus den Referenzstrukturen. Daher können in diesem Fall Ergebnisse mit und ohne kristallisiertem Wasser miteinander verglichen werden. Es werden außerdem vier Mutationen an der Referenzstruktur ESO_3003 durchgeführt. Dies sind die TOP_3 Mutationen, einzeln und gemeinsam. Diese Mutationen zeigen eine größere Affinität zu Xylose als die Referenzstruktur ESO_3003 und die Referenzstruktur mit kristallisiertem Wasser, ESO_3003 H₂O. Dieses Ergebnis ist fragwürdig und ohne experimentelle Daten nicht hinreichend diskutierbar. Ein weiterer Schritt, den man in einem solchen Fall

3.1 Modellierung spezifischer GDH Isoenzyme

machen kann, ist die Erhöhung des Konformationsvolumens, z.B. durch Molekulardynamiksimulationen. Dies erfordert jedoch einen extrem höheren Rechenaufwand. Eine Erhöhung der Umsetzungsgeschwindigkeiten bei den Mutationen ESO_3024 K110R und ESO_3024 V134E würden hingegen sinnvoll sein, da sie, wie oben diskutiert, einen direkten Einfluss auf die Monosaccharidbindung haben. Die Pfortnerfunktion findet bei der ESO_3024 R70L Mutation beim Docking keine Beachtung. Das Programm kann den Liganden willkürlich positionieren (siehe Kapitel 2.2). Da diese Aminosäure aber vermutlich nicht direkt am Monosaccharid bindet, ist diese Bewertung einer schwächeren Affinität, vermutlich dem Ladungsverlust, geschuldet.

Tabelle 3.4: Docking Ergebnisse für ESO_3003 und ESO_3024. Der Zusatz H₂O zeigt die Berücksichtigung von kristallisiertem Wasser an. TOP_3 bezieht sich auf die in Abbildung 3.5 gefundenen Punktmutationen. Die Werte für die relative Umsetzungsgeschwindigkeiten für ESO_3003 basieren auf Arbeiten von Lamble et al. [15]. Informationen für die relative Umsetzungsgeschwindigkeiten für ESO_3024 wurden in persönlichen Gesprächen mit Dr. Eleni Theodosiou erhalten. Die Bindungsenergien von AutoDock wurden als *kcal/mol* ausgegeben.

Kandidat	Ligand	Bindungsenergie	Relativer Umsatz
ESO_3003	Xylose	-5,87	n.a.
ESO_3003 H ₂ O	Xylose	-5,83	8
ESO_3003 L50R	Xylose	-6,14	n.a.
ESO_3003 R90K	Xylose	-6,93	n.a.
ESO_3003 E114V	Xylose	-6,39	n.a.
ESO_3003 TOP_3	Xylose	-6,5	n.a.
ESO_3003 H ₂ O	Glucose	-6,53	100
ESO_3003 H ₂ O	6-deoxy-Glucose	-6,29	57
ESO_3003 H ₂ O	2-deoxy-Glucose	-5,71	1
ESO_3024	Xylose	-6,17	0
ESO_3024 R70L	Xylose	-4,63	0
ESO_3024 K110R	Xylose	-6,63	n.a.
ESO_3024 V134E	Xylose	-6,68	n.a.
ESO_3024 TOP_3	Xylose	-5,96	n.a.

3.2 Modellierung von Inhibitorpositionen in Chitinase B

3.2.1 Einführung

Für eine kompetitive Hemmung von Chitinasen sind klassenspezifische Inhibitoren, wie Chitobiose oder Allosamidin (Abschnitt 1.2.2), bekannt. Optimierte für Chitinasen spezifische Peptidinhibitoren können die enzymatische Aktivität ebenfalls kompetitiv hemmen. Diese Peptide besitzen, im Gegensatz zu herkömmlichen Inhibitoren, eine höhere Spezifität gegenüber einzelnen Chitinasen. In der Zukunft könnten hierdurch, für den Menschen pathogen wirkende Chitinasen gehemmt werden, ohne die Funktion der menschlichen Chitinase zu beeinflussen.

Im Jahr 2008 führten Petter et al. [138] ein Phagen-Display-Screening durch und fanden stark bindende Peptid-Inhibitoren. Sie beschrieben diese hochspezifischen Inhibitoren für die Chitinase A und B von *Serratia marcescens*. Tabelle 3.5 zeigt eine Auswahl dieser Arbeit. Im folgenden werden die Systeme nur durch die Namen ihrer Inhibitoren benannt.

Tabelle 3.5 zeigt, dass für Chitinase A nur ein Peptidinhibitor gefunden wurde. Darüber hinaus ist die Affinität dieses Inhibitors deutlich geringer als die Affinitäten der für die Chitinase B gefunden Inhibitoren. Daher sah Wolfgang Höhne (Leiter der damaligen Forschungsgruppe an der Charité Berlin (Institut für Biochemie, Abteilung für Proteinstrukturforschung)) Möglichkeiten zur Verbesserungen und bot Daniel Hoffmann (Leiter dieser Arbeitsgruppe) eine Kooperation zur Findung von Punktmutationen für ICA3 an. Wir erhielten hierzu von Norbert Krauss, damaliges Mitglied der Arbeitsgruppe von Wolfgang Höhne und heute Dozent an der Queen Mary Universität (London), nicht publizierte gelöste Strukturen von

Tabelle 3.5: K_i Werte aus der Arbeit von Petter et al. [138]. Die Buchstaben in Fettschrift zeigen die Sequenzunterschiede in den Peptidinhibitoren.

Enzym	Peptidname	Peptidsequenz	$K_i \times 10^6 (M^{-1})$
Chitinase A	ICA3	ACPWRAPGLCG	8,1
Chitinase B	ICB1	ACR H CDS P HGG	0,54
Chitinase B	ICB1A	ACR M CDS P HGG	0,051
Chitinase B	ICB1B	ACR G CDS P HGG	0,041
Chitinase B	ICB1a	ACR M CDS P H W G	0,0045
Chitinase B	ICB1b	ACR G CDS P H W G	0,0037

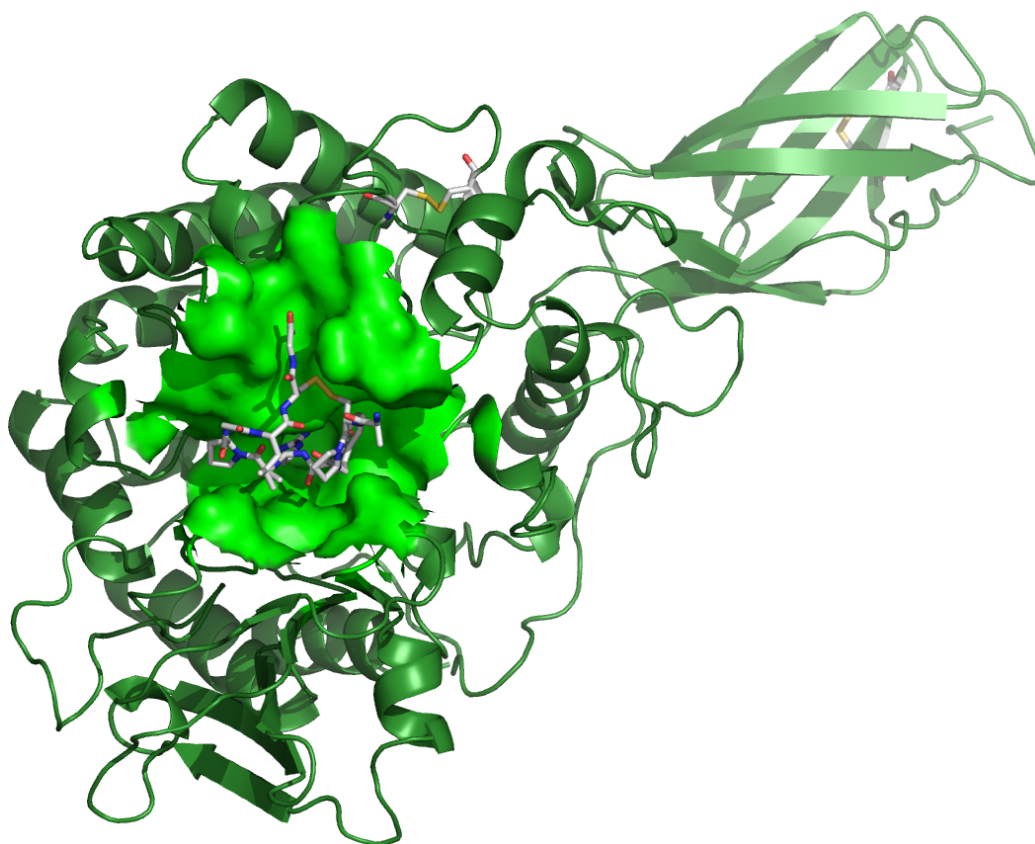


Abbildung 3.7: Chitinase A mit Peptidinhibitor ICA3. In der Struktur sind vier Cystein Aminosäuren zusätzlich eingezeichnet, die zwei Disulfidbrückenbindungen ausbilden.

ICA3 (Abbildung 3.7) und ICB1 (Abbildung 3.8) im PDB-Format.

Basierend auf Ergebnissen meiner Diplomarbeit [139], und Publikationen über den erfolgreichen Einsatz des Dockings für Peptidliganden [140, 141, 142, 143], soll im Folgenden untersucht werden, wie vielversprechend neueste Techniken des Dockings in diesem speziellen Fall sind, um repräsentative Strukturmodelle für die obige Fragestellung zu erhalten.

In dieser Arbeit werden zwei dieser neuesten Techniken angewendet. Zuerst wird ein sogenanntes *cross*-Docking von bekannten Inhibitoren für die Chitinase B durchgeführt. Teil dieses *cross*-Dockings ist das sogenannte *re*-Docking. Es hilft, das System und die Ergebnisse eines späteren Dockings mit unbekannten Liganden besser einschätzen zu können. Ein *cross*-Docking liefert darüber hinaus

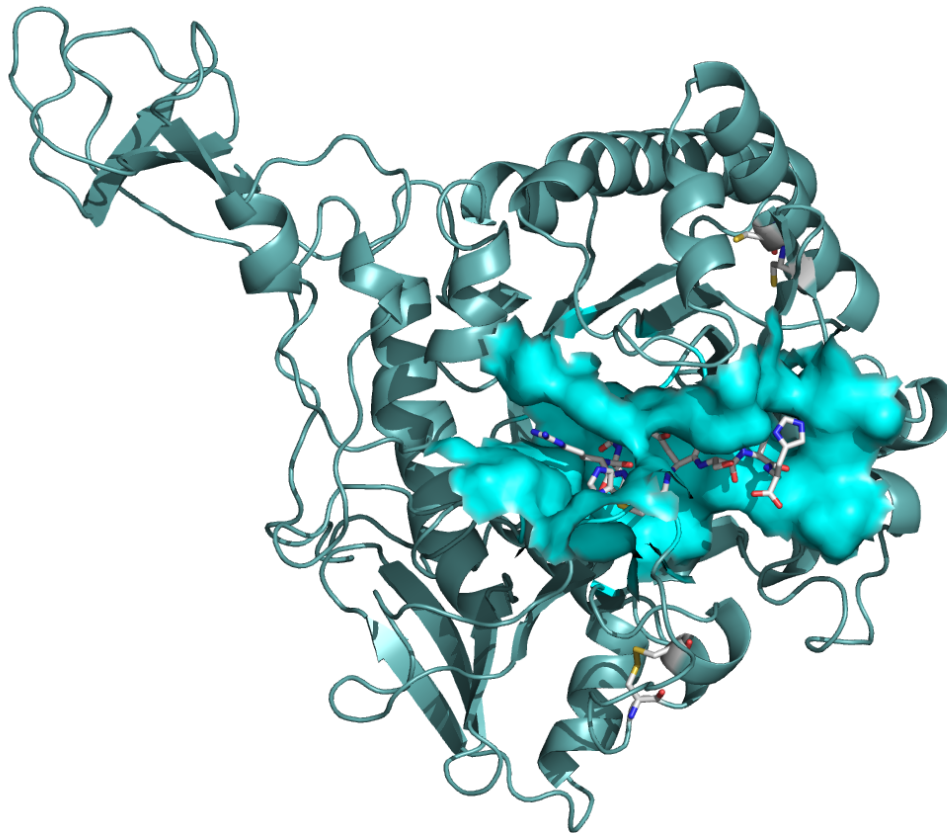


Abbildung 3.8: Chitinase B mit Peptidinhibitor ICB1. Die Struktur zeigt vier Cystein Aminosäuren, davon gehen aber nur zwei eine Disulfidbrückenbindung ein.

Einblicke in den Induced-Fit Effekt des Systems. Zweitens wird der Einfluss eines sogenannten *re*-Scorings mit dem MMGBSA Verfahren, in Zusammenhang mit ICB1 und seinen Punktmutationen, untersucht. Im Vergleich zum klassischen MMGBSA Verfahren (Kapitel 2.6.2) wird es in einem Dockingansatz leicht modifiziert verwendet. Nichts desto trotz konnte mittlerweile gezeigt werden, dass dies eine nützliche Maßnahme zur Erweiterung des Dockings ist [144, 145, 142, 146, 147, 148, 149, 150, 151, 152, 153, 41, 154].

3.2.2 Materialien und Methoden

PDB Strukturen Die von Norbert Krauss erstellten PDB-Dateien (Kapitel 3.2.1) enthalten einmal die Struktur ICA3 und zum anderen die Struktur ICB1. Es fällt

auf, dass der Inhibitor ICB1 nicht vollständig aufgeklärt ist. Es fehlten die letzten beiden Aminosäuren des Inhibitors. Bei beiden Strukturen startet die Zählung der Aminosäuren nicht mit 1, sondern mit 24 bzw. 3. Beide Dateien enthalten keine zusätzlichen Informationen im Datei-Kopf. Somit ist die Qualität der Strukturen nicht näher zu bestimmen. Des Weiteren sind in Chitinase A zwei Disulfidbrückenbindungen zu finden. In Chitinase B befinden sich die C_{α} -Distanzen von zwei anderen Cysteinseitenketten in einer ähnlichen Entfernung, wie bei der geschlossenen Disulfidbrückenbindung. Hier sind die Seitenketten aber nicht zueinander ausgerichtet.

Durchführung von *re*- und *cross*-Docking Es wird ein *re*-Docking im Rahmen des *cross*-Docking Ansatzes durchgeführt. Im *cross*-Docking Ansatz wird neben 15 Chitinase B Strukturen von *Serratia marcescens* (Tabelle 3.6), auch die von Norbert Krauss zur Verfügung gestellte Chitinase B Struktur verwendet. Die 15 Strukturen wurden auf der Grundlage einer NCBI BLAST [155] Suche, basierend auf PDB Strukturen als Suchkriterien, ausgewählt.

Alle 15 Dimere werden in ihre Ketten aufgeteilt, so dass 30 Monomere vorliegen. Alles außer den schweren Atomen der Monomerketten und deren Inhibitoren (Tabelle 3.6) wird aus den Dateien gelöscht, einschließlich aller Wasserstoffatome. Die 30 Ketten werden strukturell, mit PyMOL [129] auf die Chitinase von Norbert Krauss ausgerichtet. Alle Liganden werden in dieser ausgerichteten Position extrahiert und separat gespeichert. Jeder Ligand wird in allen 31 Ketten 20 mal für 20 verschiedene Konfigurationen mit AutoDock Vina [39] gedockt. Das Docking wird global durchgeführt, wobei auf der gesamten Chitinaseoberfläche nach möglichen Bindungspositionen gesucht wird. Vorbereitende AutoDock Vina Skripte werden für die Herstellung der benötigten PDBQR-Dateien verwendet. Die gedockten Positionen werden mit den Positionen in den entsprechenden PDB Strukturen verglichen. Für diesen Vergleich wird *g_rms*, ein Werkzeug von GROMACS [165], verwendet. Die Option *mass-weighting* wird verwendet, außerdem wird nicht zur Referenzstruktur gefittet.

Re-Scoring Das Peptid ICB1 wird nicht vollständig in der zur Verfügung gestellten Struktur gezeigt. Hier fehlten die letzten beiden Aminosäuren, Glycin und

3.2 Modellierung von Inhibitorpositionen in Chitinase B

Glycin. Diese beiden Aminosäuren und ihre Punkt-Mutationen, in ICB1a, ICB1b, ICB1A und ICB1B, werden manuell mit PyMOL 1.2r2 [129] und der *Wizard*-Funktion Mutagenese durchgeführt. Alle fünf Peptide werden mit einem globalen Docking an die Chitinase B von Norbert Krauss gedockt. Hierbei wird jeder Ligand 20 mal für 20 verschiedene Konfigurationen mit AutoDock Vina [39] gedockt. Vorbereitende Skripte von AutoDock Vina werden für die Herstellung der benötigten PDBQR-Dateien verwendet. Alle Posen innerhalb der Bindungstasche werden für ein *re*-Scoring mit dem MMGBSA Verfahren genutzt. Dazu werden die Chitinase-Inhibitor-Komplexe in eine dodekaedrische Box, mit TIP3P Wasser und einem Abstand zu den Kanten von 12 Å und einer Nähe von 0,8 Å zum Komplex, getaucht. Die Box wird mit Natrium- und Chloridionen neutralisiert. Das

Tabelle 3.6: In dieser Arbeit verwendete PDB Strukturen der Chitinase B von *Serratia marcescens*. Das jeweilige Taschenvolumen wurde mit PocketAnalyzerPCA [156] berechnet. Die vollständigen Namen der Liganden werden im *Abkürzungsverzeichnis* dieser Arbeit gezeigt. NAG lag als Pentamer vor [157, PDB ID: 1e6n]. AMI,NAA lag als NAA-NAA-AMI Polymer vor [157, PDB ID: 1e6r] [158, PDB ID: 1ogg]. NAG,NGO lag in Kette A als NAG-Dimer und NGO, und in Kette B als NAG-Monomer und NGO, vor.

PDB	Mutant	Volumen Bindetasche		Ligand IDs	Lit.
		Kette A	Kette B		
1w1p	–	1784	1460	GIO	[159]
1w1v	–	2181	2020	ALJ	[159]
1w1y	–	1918	1993	TYP	[159]
1ogb	D142N	1700	1303	–	[158]
1ogg	D142N	1507	1489	AMI,NAA	[158]
1ur8	–	1604	1513	GDL,NAG	[160]
1ur9	–	1797	1694	GDL,NAG,PHJ	[160]
1o6i	–	1551	1370	0HZ	[161]
1gpf	–	1660	1581	–	[162]
1goi	D140N	1637	1482	–	[163]
1e6n	E144Q	2037	1914	NAG	[157]
1e6p	E144Q	1650	1463	–	[157]
1e6r	–	1693	1417	AMI,NAA	[157]
1e6z	–	1813	1517	NAG,NGO	[157]
1e15	–	1659	1555	–	[164]
Chitinase B	–	1745	–	–	

System wird unter Verwendung des AMBER03 Kraftfeldes für Proteine [166] parametrisiert. Eine erste Minimierung wird durchgeführt, um Wassermoleküle zu minimieren. Daher werden alle schweren Atome des Komplexes mit einer Kraftkonstante von $4\text{kcal mol}^{-1}\text{\AA}^{-2}$ in ihrer Bewegung eingeschränkt. Die Minimierung wird für 100 Schritte *steepest descent*, gefolgt von 100 Schritte *conjugate gradient* durchgeführt. Nicht gebundene Wechselwirkungen werden bis zu einer Distanz von 9\AA , in einem NVT Ensemble, berechnet. Eine daran angeschlossene Minimierung nutzt die gleichen Parameter, wobei nun das System nicht mehr beschränkt wird, um alle schweren Atome des Komplexes zu minimieren. Diese Minimierung wird für 1000 Schritte *steepest descent*, gefolgt von 1000 Schritte *conjugate gradient* durchgeführt. Der letzte Zeitschritt der zweiten Minimierung wird für die MMGBSA Methode 2.6.2 verwendet. MMPBSA.py [167] wird mit seinen Standardwerten verwendet. Die elektrostatische Solvatationsenergie wird unter Verwendung des *Generalized Born Models I*, von Onufriev et al. [79] berechnet.

3.2.3 Ergebnisse und Diskussion

Cross-Docking und re-Docking Ein *cross*-Docking ist nützlich, um mehr über das System zu erfahren und die Anwendbarkeit des Dockings auf das System zu verifizieren. Darüber hinaus kann das Verfahren damit auf die Systemanforderungen abgestimmt werden. In dieser Arbeit ist das *re*-Docking Teil des *cross*-Dockingansatzes. Die Ergebnisse des *cross*-Dockings sind in Abbildung 3.9 und 3.10 gezeigt. Die RMSD wird hier aus allen schweren Atomen eines Liganden berechnet. Als Referenzposition wird die Position aus der jeweiligen PDB gewählt, aus der ein Ligand extrahiert wird. Da die meisten Rezeptoren der Chitinasen Gemeinsamkeiten aufweisen, werden die RMSD Ergebnisse in einem Box-Plot gezeigt, basierend auf den Eigenschaften von 10 Rezeptorgruppen (Abbildung 3.9). Die erste Gruppe, OWN, ist das *re*-Docking und bedeutet, dass hier nur Chitinasen mit ihren Liganden aus den jeweiligen PDB-Dateien berücksichtigt werden. ALL zeigt das Ergebnis für alle Liganden mit allen Rezeptoren. Hierbei wird das Docking aller Liganden zusätzlich noch in APO- und HOLO-Strukturen aufgeschlüsselt. Einige Kristallstrukturen haben Punktmutationen in der Bindungsta-

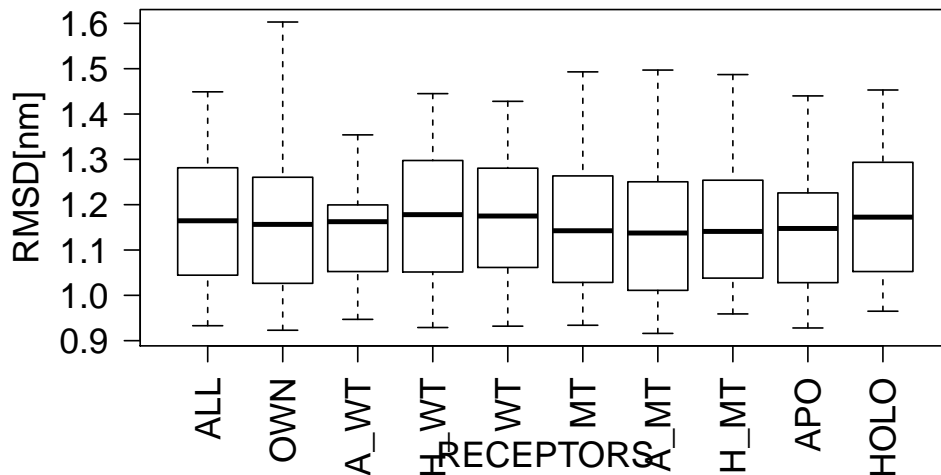


Abbildung 3.9: Cross-Docking Ergebnisse nach Chitinasen Aufgeteilt nach RMSD gegen Rezeptorgruppen, wobei eine Beschreibung der Gruppen im Text zu finden ist.

sche. Daher sind die Strukturen außerdem nach Strukturen mit Punktmutationen (Mutanten; MT) und ohne Punktmutationen (Wildtyp; WT) gruppiert. Schließlich ist es auch möglich, dass die APO- und HOLO-Ergebnisse durch Punktmutationen beeinflussbar sind. Aus diesem Grund gibt es noch eine Unterteilung A_WT, H_WT, A_MT und H_MT. Während sich A hierbei auf APO bezieht, ist H die Abkürzung für HOLO.

Die erste Auffälligkeit in dieser Abbildung ist die Ähnlichkeit des Median in allen Gruppen. Dies zeigt, dass sich alle ausgewählten Gruppen mehr oder weniger gleichen und dass keine klare Vorliebe für die jeweils gelöste Rezeptorkonfigurationen (OWN), was eigentlich zu erwarten wäre, vorliegt. Stattdessen zeigt sich hier, über alle Gruppen hinweg, die größte Differenz zwischen dem kleinsten und größten Wert. Ein Nachteil dieser Art der Analyse ist die ungleiche Größe der Datensätze in jeder Gruppe. Dennoch zeigt gerade dieser Umstand den besonderen Einfluss der APO- und MT-Strukturen. Die MT-Strukturen und ihre Variationen (A_MT, H_MT) zeigen die geringsten Medianwerte. Interessanterweise zeigen die APO-Strukturen in allen Fällen bessere Ergebnisse als die HOLO-Strukturen. Dies ist

3.2 Modellierung von Inhibitorpositionen in Chitinase B

auf den ersten Blick eine Überraschung. In den Abbildungen 3.7 und 3.8 ist gezeigt, dass die Bindungstaschen in den Chitinasen vergraben sind. Aus Tabelle 3.6 ist bekannt, dass die APO-Strukturen nicht das größte Volumen der Bindetaschen besitzen. Die Abweichung der APO-Ergebnisse in Wildtyp (A_WT) und Mutanten (A_MT) zeigt, dass der A_WT Datensatz das beste Gesamtergebnis liefert. Der in Abbildung 3.10 dargestellte Boxplot präsentiert die Ergebnisse nach Liganden aufgeteilt. Es ist bekannt, dass der Suchraum im Docking mit den Freiheitsgraden eines Liganden korreliert (Abschnitt 2.2). In unserem Fall korrelieren die Freiheitsgrade mit der Menge der Atome (Tabelle 3.7). Eine einzige Ausnahme zeigen der Ligand NAG-NAG-NAG-NAG-NAG und das Peptid ICB1. Aus diesem Grund zeigt die x-Achse nicht die Namen der Liganden, sondern die Menge der schweren Atome eines Liganden, welche sich von links nach rechts erhöhen. Aus diesem Plot kann abgeleitet werden, dass das Auffinden der exakten Bindungsposition nicht direkt mit der Größe bzw. der Menge der frei drehbaren Bindungen korreliert. Stattdessen zeigt der Plot verschiedene RMSD Werte der beiden Liganden mit 18 schweren Atomen. Im Gegensatz dazu weisen die bei-

Tabelle 3.7: Gedockte Liganden der Chitinase B. Die Menge der schweren Atome (ATOME) sowie der frei drehbaren Bindungen (BINDUNGEN) ist dargestellt. Mit nur einer Ausnahme, korreliert die Menge an schweren Atomen mit der Menge der frei drehbaren Bindungen. Nur bei NAG-NAG-NAG-NAG-NAG und ICB1 trifft diese Regel nicht zu.

Ligand	BINDUNGEN	ATOME
GIO	0	11
NGO	1	14
NAG	2	15
TYP	2	19
ALJ	4	18
OHZ	4	18
NAG-GDL	6	29
NAG-NAG	6	29
NAG-GDL-PHJ	9	39
NAA-NAA-AMI	10	43
NAG-NAG-NAG-NAG-NAG	18	71
ICB1	25	49

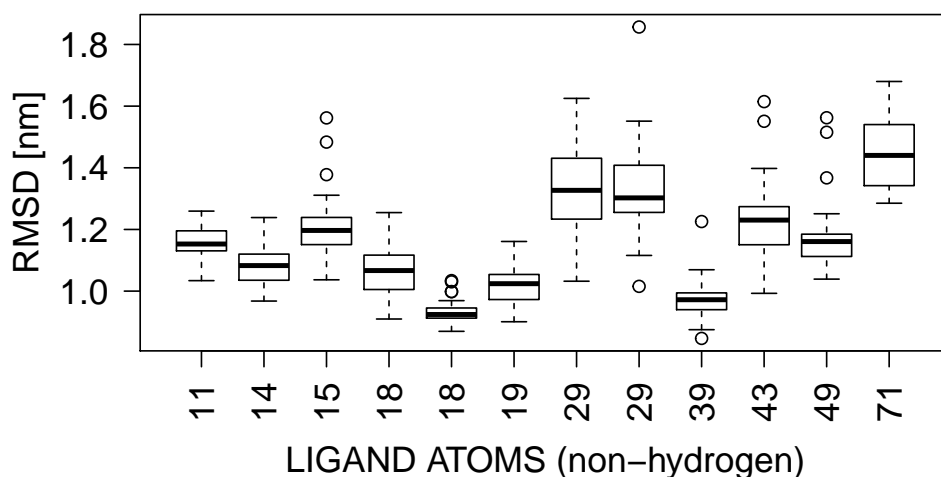


Abbildung 3.10: Cross-Docking Ergebnisse nach Liganden Aufgeteilt nach RMSD gegen die Menge der schweren Atome eines Liganden. Eine Aufschlüsselung der schweren Atome nach ihren Liganden ist in Tabelle 3.7 zu finden.

den Liganden mit 29 schweren Atomen die gleichen RMSD Werte auf. Sowohl im Falle der 18 schweren Atome, als auch im Falle der 29 schweren Atome werden sehr ähnliche Strukturen miteinander verglichen. Im Fall der 18 schweren Atome wird *cyclo*-(L-Arg-D-Pro) mit *cyclo*-(L-Arg-L-Pro) verglichen. *Cyclo*-(L-Arg-L-Pro) zeigt im Mittel einen kleineren RMSD Wert. Im Fall der 29 schweren Atome stellt sich der Unterschied in den Strukturen nur durch eine Carbonyl- gegen eine Hydroxygruppe dar. Dies unterstreicht eine nicht lineare Korrelation des richtigen Bindungsmodus mit der Ligandengröße. Peptid ICB1 ist in der Abbildung mit dem Index 49 zu finden. Das gefundene Ergebnis kann dabei als stabil mit guter Nähe zur Referenzposition beschrieben werden, da die Quantile und die Minimum- und Maximumwerte in der näheren Umgebung des Medians zu finden sind. Ausreißer, wie in diesem Fall von ICB1, sind wahrscheinlich leicht zu identifizieren, da sie nicht oder nur teilweise in der Bindungstasche angeordnet sind. Es ist nun möglich, die gefundenen Ergebnisse in den Strukturen zu diskutieren. Leider ist dies im zeitlichen Rahmen dieser Arbeit nicht mehr möglich.

3.2.4 Re-Scoring

Zum Zeitpunkt dieser Untersuchungen wird das MMPBSA bzw. MMGBSA Verfahren vielversprechend in anderen Ansätzen, wie z.B. dem BEAR-Ansatz [150], implementiert oder intensiv als *re*-Scoring Verfahren für Docking Ansätze verwendet [146]. Daher liegt es nahe, diesen Ansatz auch für ICB1 und seine punktmutierten Peptide zu testen. Meine Berechnungen ähneln dabei der Arbeit von Hou et al. [146], die heraus fanden, dass die GB Berechnungen in einem Docking Ansatz PB Berechnungen übertreffen und Rechnungen auf einer einzigen minimierten Struktur in den meisten Fälle ausreichend sind. In den Arbeiten von Hou et al. wurde der entropische Term vernachlässigt, in dieser Arbeit wird er zusätzlich untersucht. Tabelle 3.8 zeigt die Ergebnisse dazu. Die Spalten sind nach absteigenden experimentellen Affinitäten sortiert. Des Weiteren ist die Tabelle in drei horizontale Blöcke aufgeteilt. Der erste Block zeigt die AutoDock Vina Er-

Tabelle 3.8: Re-Scoring der Affinitäten von AutoDock Vina und MMGBSA. Alle Ergebnisse sind in *kcal/mol* gezeigt. TOP 1, TOP 3 und TOP 5 beziehen sich auf den Mittelwert der Ergebnisse mit den jeweils stärksten Affinitäten. NUR NEG. WERTE berücksichtigt nur Entropien mit negativem Vorzeichen.

	ICB1b	ICB1a	ICB1B	ICB1A	ICB1
AutoDock Vina					
ALLE ERGEBNISSE	-9,5	-9,3	-8,8	-8,4	-9,1
TOP 1	-11,3	-11,5	-10,7	-10,3	-10,9
TOP 3	-11,2	-11,5	-10,5	-10,3	-10,8
TOP 5	-11,2	-11,4	-10,4	-10,2	-10,8
MMGBSA (ΔH)					
ALLE ERGEBNISSE	-67,0	-72,1	-63,1	-68,2	-68,8
TOP 1	-135,0	-142,7	-121,3	-130,5	-127,6
TOP 3	-132,4	-139,5	-120,7	-127,2	-126,3
TOP 5	-130,4	-137,3	-119,2	-125,8	-123,9
MMGBSA ($\Delta H - T\Delta S$)					
NUR NEG. WERTE	-28,5	-31,3	-33,3	-31,7	-29,2
ALLE ERGEBNISSE	17,6	11,1	10,8	9,4	10,9
TOP 1	-99,4	-100,1	-112,7	-118,5	-100,5
TOP 3	-88,9	-97,9	-105,1	-107,2	-97,8
TOP 5	-84,0	-95,8	-98,1	-103,2	-94,5

gebnisse. Sie sind dabei mit Ausnahme von ICB1 in der richtigen Reihenfolge. Leider zeigt sich dieses Ergebnis nicht, wenn nur ein, drei oder fünf Docking Posen verwendet werden. Zur Erreichung dieser Ergebnisse ist ein großes Ensemble an Posen, hier 400 pro Peptid, notwendig. Es kann also nicht auf einfachem Wege eine Lösung zur Beschreibung der atomaren Wechselwirkungen erreicht werden, da zusätzliche Schritte zur Selektierung individueller Strukturen nachgeschaltet werden müssen. Anderen Ergebnissen zufolge, wie bei Gouda et al. [142], ist ein *re*-Scoring mit mehreren Bewertungsfunktionen vielversprechend. Methoden wie *Rank-by-Rank* oder *Rank-by-Number* [168] und die AASS (Average of Auto-Scaled Scores) [169] können genutzt werden, um ein Gesamtscore zu erhalten. Der nächste horizontale Block zeigt die MMGBSA Ergebnisse nur auf Grundlage des entropischen Terms ΔH . Auf den ersten Blick sind die Ergebnisse hierbei willkürlich und ohne Trend. Auch die Aufnahme des Entropieterms $-T\Delta S$ führt nicht zur experimentell gefundenen Reihenfolge. Dennoch können beide Berechnungen zwischen einem Glycin und einem Tryptophan an der zehnten Peptidposition unterscheiden (Tabelle 3.5). Das ursprüngliche Problem, ein korrektes Abbilden der experimentellen Reihenfolge, wird trotzdem nicht durch einen MMGBSA Ansatz gelöst. Dies gilt sowohl für das MMGBSA als auch für das AutoDock Vina Scoring, wobei das Problem ähnlich gelagert zu sein scheint. Entweder überschätzen beide Methoden mit ihren Parametern Histidin, an ICB1 Position 4, oder das Docking selbst kann keine repräsentativen Ensembles für die vier punktmutierten Liganden vorschlagen. Wenn man die besten Posen miteinander vergleicht, ist es offensichtlich, dass sie nicht gleich sind. Beispielsweise werden nur drei der AutoDock Vina TOP 5 Posen durch den MMGBSA Ansatz reproduziert. Darüber hinaus sind auch die besten ΔH Posen verschieden von den besten $-T\Delta S$ Posen. Aus diesem Grund erscheint auch der Ansatz von Gouda et al. [142] als sinnvoll, wo ein *re*-Scoring nur für die besten Posen des Dockings durchgeführt wurde. Tabelle 3.9 zeigt die Ergebnisse dieser Vorgehensweise für die TOP 1, TOP 3, TOP 5 und TOP 1% der Docking Posen von AutoDock Vina. In diesem Fall sind die ΔH Ergebnisse der TOP 1% und TOP 5 Posen von ICB1 richtig sortiert. Leider zeigt diese Art der Bewertung Probleme mit dem Scoring von ICB1a und ICB1b. Strukturell ist hier der Unterschied in einem Methionin, in ICB1a, und einem Glycin, in ICB1b, zu finden.

Durch diese Experimente können Ansätze gezeigt werden, um Docking Experimente auf ein spezielles System zu zuschneiden. Ein Mangel an Zeit verhindert leider eine Auslotung und vollständige Durchführung aller Möglichkeiten, um bessere Methodiken zu entwickeln und die speziellen Eigenschaften des Systems noch klarer darzustellen. Vielversprechende Schritte wären an dieser Stelle, kurze Simulation der minimierten Strukturen [142, 146] oder ein wiederholtes Docking mit den minimierten Strukturen [143] durchzuführen. Auch die Verwendung von APO Rezeptoren ist denkbar (siehe Kapitel 3.2.3). Wie die obige Diskussion zeigt, haben alle Methoden Vor- und Nachteile, wodurch beim Ranking in beiden Tabellen an unterschiedlichen Stellen Probleme auftreten. Weitere Schritte sollten deshalb zum Ziel haben, die Vorzüge aus beiden Untersuchungen in einem Ansatz zu kombinieren. Deshalb können Methoden wie *Rank-by-Rank* zum jetzigen Zeitpunkt als vielversprechend angesehen werden.

Tabelle 3.9: MMGBSA *re*-Scoring der besten AutoDock Vina Posen. Alle Ergebnisse sind in *kcal/mol* gezeigt. TOP 1, TOP 3 und TOP 5 bezieht sich auf die AutoDock Vina Posen mit den höchsten AutoDock Vina Affinitäten. In den beiden MMGBSA Zeilen erhalten diese Posen mit MMGBSA ein *re*-Scoring. Als TOP 1% bezieht sich auf die ersten 1% der AutoDock Vina Posen mit den höchsten AutoDock Vina Affinitäten.

	ICB1b	ICB1a	ICB1B	ICB1A	ICB1
	AutoDock Vina				
TOP 1%	-11,2	-11,5	-10,4	-10,3	-10,8
TOP 1	-11,3	-11,5	-10,7	-10,3	-10,9
TOP 3	-11,2	-11,5	-10,5	-10,3	-10,8
TOP 5	-11,2	-11,4	-10,4	-10,2	-10,8
	MMGBSA (ΔH)				
TOP 1%	-82,3	-84,6	-79,9	-72,5	-63,6
TOP 1	-67,6	-62,6	-72,3	-125,8	-56,9
TOP 3	-65,4	-66,4	-82,5	-79,9	-63,4
TOP 5	-80,8	-83,4	-74,8	-69,7	-64,4
	MMGBSA ($\Delta H - T\Delta S$)				
TOP 1%	-102,3	-105,1	-83,3	-71,8	-85,0
TOP 1	-70,2	-63,3	-121,2	-103,2	-55,2
TOP 3	-90,7	-92,4	-83,7	-96,0	-89,5
TOP 5	-101,3	-104,6	-78,4	-68,7	-88,9

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

3.3.1 Einführung

Im Jahre 1977 wurde die erste Molekulardynamiksimulation (Kapitel 2.4) eines Proteins durchgeführt [170]. Mittlerweile gehören neben dem Docking zur Abtastung von Energielandschaften Simulationen zum Standardwerkzeug der biomolekularen Modellierung [7]. Auch mit den heute verfügbaren Rechnern ist das Abtasten des vollen Konformationsraums eines Proteins oder eines Protein-Ligand-Komplexes oft noch immer nicht durchführbar. Daher ist es durchaus üblich, eine relativ kleine Teilmenge der Konformationen in geometrischer Nähe zur Ausgangsstruktur zu simulieren. Typischerweise wird hierbei in Größenordnungen von 10 – 100 *ns* gearbeitet. Oft müssen vor solchen Simulationen die verwendeten Strukturen zunächst generiert werden, beispielsweise durch eine vergleichende Modellierung (Kapitel 2.1), da sie in strukturellen Datenbanken wie der PDB [46] nicht verfügbar sind. In solchen Fällen drängt sich die Frage auf, ob die in den Simulationen generierten Konformationen repräsentativ für die Systeme sind. Eine Möglichkeit, dies in einer sehr einfacher Art und Weise zu analysieren, ist die Einführung einer Störung in eine bekannte Struktur. Beispielsweise kann die Struktur einer Seitenkette durch Rotation verändert werden (Abbildung 3.11). Eine Simulationen der gestörten Struktur (Modellstruktur) und der bekannten Struktur (Referenzstruktur) sollte anschließend in einem zeitlich angemessenen Rahmen einen gleichen Konformationsraum aufspannen.

Dieses Kapitel ist eine Fallstudie auf der Grundlage der oben beschriebenen Zusammenhänge. Hierzu wurde die Hck-SH3 Domäne im Komplex mit dem Peptid PD1 verwendet [37, PDB ID: 2oi3]. Für eine bessere Unterscheidung zwischen SH3 Domäne und Peptid wird im Folgenden der Ein-Buchstaben-Code für die Aminosäuren der SH3 Domäne verwendet. Für die Aminosäuren des Peptids wird hingegen der Drei-Buchstaben-Code verwendet. Untersucht wird hier das Verhalten der Seitenkette von Tyr4 wie in der PDB hinterlegt und einem modellierten Rotamer. Wie in Abbildung 3.11 zu erkennen, unterscheiden sich beide Konformationen ausschließlich durch den Winkel. Der Diederwinkel $C-C_{\alpha}-C_{\beta}-C_{\gamma}$ wird dabei um etwa 100° verschoben und im folgenden mit $C-C_{\gamma}$ abgekürzt.

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

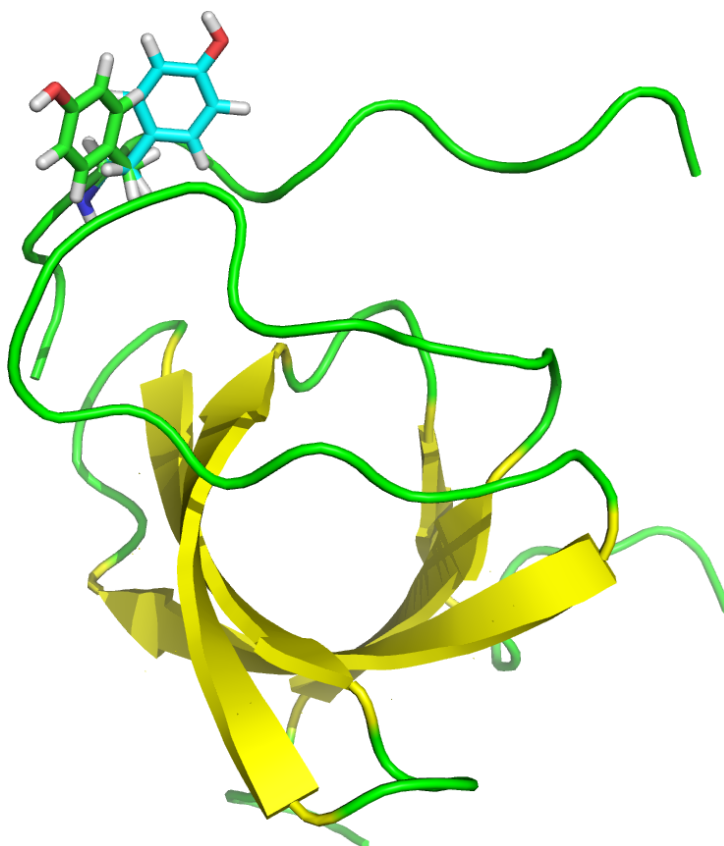


Abbildung 3.11: Unterschied in den Startstrukturen von PD1:Hck-SH3. Unterschied in den Startstrukturen C (grün) [37, 2oi3] und M (cyan) des modellierten Rotamers.

Eine Gesamtsimulationszeit von $3,2\mu s$ der beiden Strukturen (PDB (C) und Rotamer/Modell (M)) zeigt eine unerwartete Konformation des Tyr4, welche mit MMGBSA [98] und SAM [118] analysiert wird. Die Analysen zeigen, dass das Tyrosin der beiden Startstrukturen in ihren jeweiligen Simulationen nicht gleich oft die selben Konformationen einnehmen. Interessanterweise zeigen sich aber sowohl für den Diederwinkel als auch für die Strukturen statistisch nicht unterscheidbare Simulationen. Heutige Simulationsansätze mit mehreren Trajektorien werden oft in einer Makrotrajektorie analysiert [171]. Dabei werden die einzelnen Simulationen aneinander gehängt. Die Analyse von Makrotrajektorien nimmt aber implizit an, dass die zu analysierenden Zeitschritte unabhängig voneinander

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

sind. Dies ist aber nicht immer der Fall [90]. Mit einer Signifikanzanalyse, wie SAM (Kapitel 2.6.5), wird dieses Problem vermieden.

3.3.2 Material und Methoden

Strukturvorbereitung Alle Schritte zur Strukturvorbereitung werden mit Pymol 1.2r2 [129] durchgeführt. Dazu wird die PDB Struktur von PD1:Hck-SH3 [37, PDB ID: 2oi3] verwendet. Die ersten 21 Aminosäuren der SH3 Domäne werden gelöscht, da sie nach Larson et al. [38] nicht Teil einer SH3 Domäne sind. Das Rotamer der modellierten Struktur wird auf Grundlage der in Pymol hinterlegten Rotamerbibliothek ausgewählt. Die Bibliothek enthält dabei unter Berücksichtigung des Peptidrückgrats drei Vorschläge. Es wird der Vorschlag mit den geringsten sterischen Hindernissen und der weitesten Entfernung zur ursprünglichen Struktur ausgewählt.

Simulationsprotokoll Jede Simulation wird mit der Amber11 [85] Software durchgeführt. Sander wird dabei für die Energieminimierung und PMEMD für die Simulation verwendet. Alle Strukturen werden mit leap von AmberTools 1.5 [85] vorbereitet. Dabei wird das Kraftfeld Amber99SBildn [172] verwendet. Die Protonierungszustände werden der NMR Struktur [37, PDB ID: 2oi3] entnommen. Ungeladene Histidin Seitenketten werden mit HID, entsprechende Amber Nomenklatur, in der PDB umgeschrieben. Der Komplex wird in eine oktaedrische Box mit TIP3P [173] Wasser getaucht. Dabei wird ein Abstand von mindestens 10\AA vom Komplex zu den Seiten der Box eingehalten.

Die initiale Energie der Wassermoleküle und Wasserstoffatome wird vor Simulationsbeginn mit dem *steepest descent* Algorithmus minimiert. Dabei wird ein Konvergenzkriterium von $0,0001\text{ kcal mol}^{-1}\text{\AA}^{-1}$ oder 500 Schritte voreingestellt. Zur Fixierung der schweren Atome im Protein wird ein harmonisches Potential von $100\text{ kcal mol}^{-1}\text{\AA}^{-2}$ genutzt. Die Molekulardynamiksimulationen werden in einer NPT Umgebung für 40 ns durchgeführt. Der SHAKE Algorithmus [62] wird verwendet, um eine Integrationszeit von 2 fs nutzen zu können. Die Temperatur wird bei 300 K und der Druck bei 1 bar durch den *weak coupling* Algorithmus von Berendsen [59] konstant gehalten. Beide Datensätze werden dabei jede Pi-

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

cosekunde schriftlich notiert. Die nicht gebundenen Wechselwirkungen werden bis zu einer Distanz von 8\AA , die Langstrecken-Elektrostatik wird mit dem PME Algorithmus[174] berechnet. Sowohl die C- als auch die M-Konformation werden in den 40 verschiedene Wasserboxen von Genheden et al. [90] simuliert. Dies führt initial zu unterschiedlichen Startbedingungen, welche den Verlauf jeder der 40 Simulationen verändert.

MMGBSA Protokoll MMPBSA.py [167] als Teil von AmberTools 1.5 wird mit einer Schrittweite von 10 ps , 100 ps und 1000 ps verwendet. Eine Schrittweite von 10 ps wird für die Berechnungen des Winkels C-C $_{\gamma}$ der Tyr4 Struktur verwendet. Der Solvationsterm wird mit GB1, entsprechende Tool Nomenklatur, analysiert [175, 78]. Entropische Berechnungen werden vernachlässigt. Ansonsten wird MMPBSA.py mit seinen Standardwerten verwendet. Die Daten werden mit R [176] nachbearbeitet.

SAM Protokoll Die SAM Analysen werden für C $_{\alpha}$ -, OH-, NO und apolare C-Distanzen durchgeführt. Hierbei definieren sich die apolaren C-Distanzen als Kohlenstoffatom-Kohlenstoffatom-Abstände der apolaren Aminosäuren Alanin, Isoleucin, Leucin, Methionin, Phenylalanin, Prolin, Tryptophan und Valin unter $4,6\text{\AA}$. Die OH- und NO-Abstände, wobei sich N, O und H auf die Atome Stickstoff, Sauerstoff und Wasserstoff beziehen, werden bis zu einer Distanz von $4,6\text{\AA}$ für alle Aminosäuren berechnet. Dabei werden die Distanzen innerhalb einer Aminosäure vernachlässigt. Die Abstände werden dabei auf Grundlage der unveränderten PDB ermittelt. C $_{\alpha}$ -Abstände werden für jedes Paar von Aminosäuren auf Grundlage der in der PDB gegebenen Struktur berechnet. Des Weiteren werden RMSF auf Basis der C $_{\alpha}$ -Atome jeder Aminosäure und der ganzen Aminosäuren berechnet. Für die Ermittlung der zu untersuchenden Atompaare wird die *hbond* Funktion von CPPTRAJ, implementiert in AmberTools 1.5 [177], verwendet. Alle Daten werden direkt aus den Trajektorien mit einer Schrittweite von 10 ps , extrahiert. RMSF Werte werden unter der Verwendung eines Zeitfensters von 5 ns , welches um jeweils eine ns weitergeschoben wird, berechnet. Aus den Datensätzen der vier Abstände wird das gleiche Zeitfenster, in Form von Mittelwerten und Varianzen, mittels R [176], berechnet. Die Datensätze werden für jedes Zeitfens-

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

ter in Form eines Arrays mit SAM ausgewertet [118]. Jede Zeile enthält dabei Daten einer Simulation und jede Spalte einen Datenpunkt.

Ein Array für eine C_α -Analyse enthält somit bei einem Komplex mit 77 Aminosäuren $(76 \cdot 77)/2 = 2926$ Spalten und $2 \cdot 40$ Simulationen = 80 Zeilen. SAM wird dabei verwendet, um signifikante Unterschiede zwischen den ersten 40 Zeilen und der letzten 40 Zeilen zu berechnen, wobei sich die zwei Gruppen auf die jeweils zwei Startrotamere beziehen. Hierzu wird das R-Paket `siggenes` [121] verwendet. Außer der FDR (false discovery rate) wird das Paket in seinen Standardeinstellungen verwendet. Die FDR ist dabei eine statistische Maßnahme, um die Menge der Falsch-Positiven Ergebnisse einschätzen zu können.

Populationsdynamik Die gekoppelte Populationsdynamik ist eine Anwendung aus der linearen Algebra und wird oft in der theoretische Biologie angewendet, beispielsweise um Viruspopulationen gegen Wirtszellpopulationen oder Insektenpopulationen basierend auf Eier, Larven und Erwachsenen zu untersuchen. Mathematisch werden solche Modelle durch ein Skalar des Zustandes $p(t)$, einer homogenen Population zum Zeitpunkt t und einem Vektor des Zustandes p einer strukturierten Bevölkerung beschrieben. Für eine Population von drei Spezies gilt damit:

$$p = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$$

Hieraus kann mittels einer Populationsmatrix α folgende Dynamik simuliert werden:

$$p(t+1) = A \cdot p(t) \quad (3.1)$$

Bei einer Population von drei Spezies, z.B. Diederwinkel, setzt sich die Matrix aus den folgenden linearen Gleichungen zusammen:

$$\begin{aligned} p_1(t+1) &= \alpha_{1,1}p_1(t) + \alpha_{1,2}p_2(t) + \alpha_{1,3}p_3(t) \\ p_2(t+1) &= \alpha_{2,1}p_1(t) + \alpha_{2,2}p_2(t) + \alpha_{2,3}p_3(t) \\ p_3(t+1) &= \alpha_{3,1}p_1(t) + \alpha_{3,2}p_2(t) + \alpha_{3,3}p_3(t) \end{aligned}$$

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

Dies kann auch durch den Ausdruck dargestellt werden:

$$\begin{pmatrix} p_1(t+1) \\ p_2(t+1) \\ p_3(t+1) \end{pmatrix} = \begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} \\ \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} \\ \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} \end{pmatrix} \cdot \begin{pmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \end{pmatrix} \quad (3.2)$$

Mittels R [176] werden für den Fall der Diederwinkel einer Startstruktur die Übergänge aller Trajektorien gezählt und in die Populationsmatrix α eingetragen. Mittels Gleichung 3.1 wird dann der Entwicklung der Diederwinkel vorhergesagt.

3.3.3 Ergebnisse und Diskussion

Strukturvorbereitung Die Modellierung der zweiten Startstruktur M von PD1:Hck-SH3 zeigt nur einen Unterschied zu Referenzstruktur C. Der Diederwinkel C-C $_{\gamma}$ von Tyr4 ist dabei um etwa 100° verschoben (Abbildung 3.11). Er beträgt 155,2° in der Referenzkonformation (Abbildung 3.12) und 58,6° in der Modellkonformation.

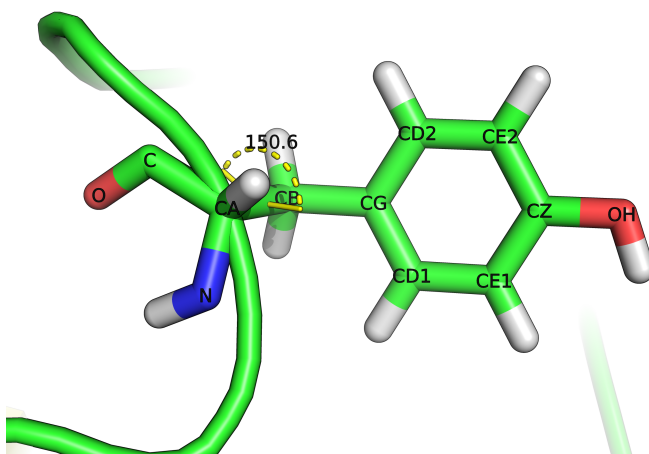


Abbildung 3.12: Nomenklatur der Atome von Tyrosin. Nahaufnahme von Tyr4 mit eingezeichnetem Diederwinkel zwischen C und C $_{\gamma}$ in Struktur C. Die schweren Atome sind dabei entsprechend der Nomenklatur in PDBs benannt, wobei die verschiedenen Kohlenstoffatome mit C, CA = C $_{\alpha}$, CB = C $_{\beta}$, CG = C $_{\gamma}$, CD = C $_{\delta}$, CE = C $_{\epsilon}$ und CZ = C $_{\zeta}$ bezeichnet werden.

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

Diederwinkel von Tyr4 Eine graphische Darstellung (Abbildung 3.13) des modellierten Tyr4 C-C γ Diederwinkels gegen die Simulationszeit zeigt drei bevorzugte Winkelkonformationen mit vielen Übergängen zwischen den Konformationen. Da eine Darstellung der Referenzstruktur dieser Abbildung stark ähnelt, wird im Folgenden nur dieser Fall diskutiert. Die Abstände liegen dabei äquivalent um jeweils 120° verschoben und sind als anti-, trans-, (+)- und (–)-gauche oder auch (+)- bzw. (–)-synklinal bekannt. Grundlage dieser Zusammenhänge sind in der Newman-Projektion zu finden, welche ausgiebig in klassischen Lehrbüchern der Organischen Chemie diskutiert werden. In dieser Abbildung entsprechen die ersten beiden Winkel, den Winkeln der Referenzstruktur bzw. modellierten Struktur. Bei der dritten Konformation ragt die Seitenkette in die SH3 Domäne hinein. Interessanterweise werden alle drei dieser Winkel als Konformation vom Mutagenese-Assistenten bei der Modellierung vorgeschlagen.

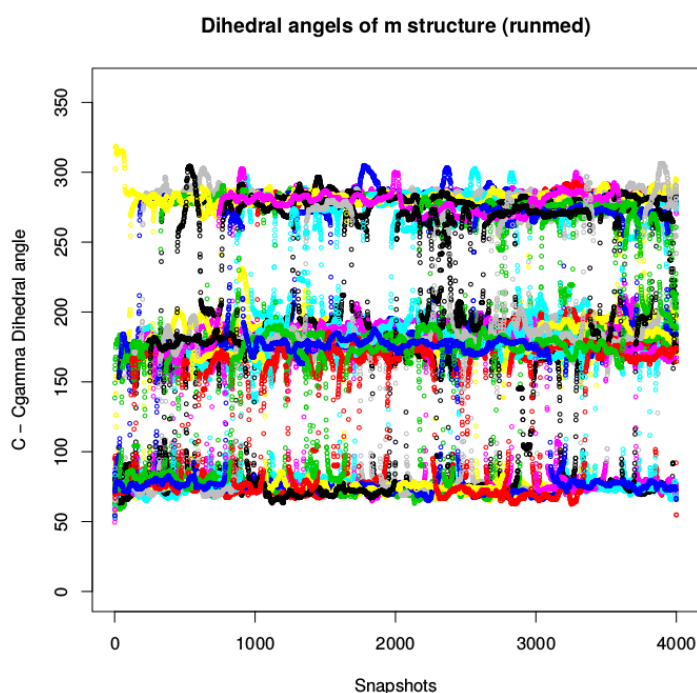


Abbildung 3.13: Tyr4 C-C γ Diederwinkel aller Simulationen. Zusammenfassung der eingenommenen Diederwinkel, mit der Modellstruktur als Startkonformation, aller Simulationen (in verschiedenen Farben), über die Zeit.

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

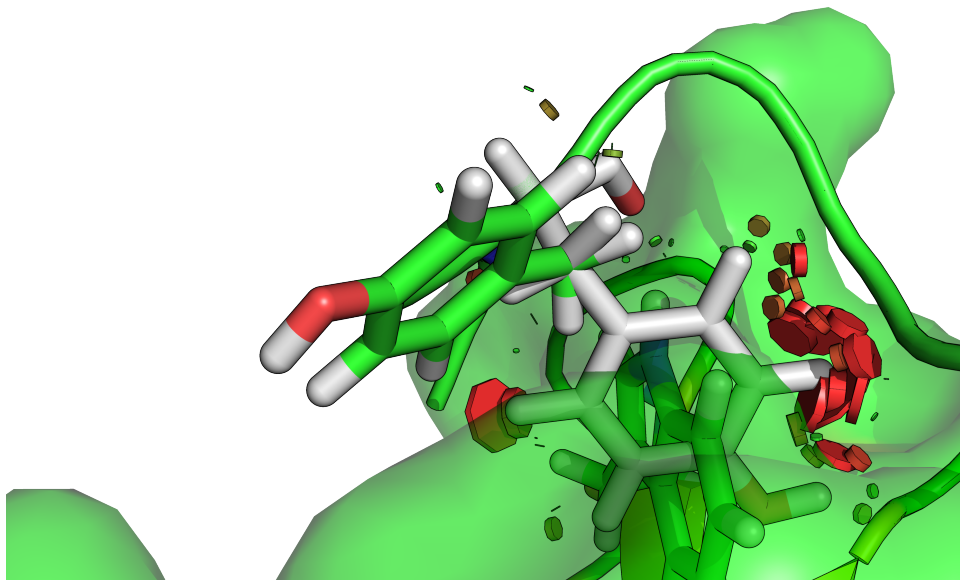


Abbildung 3.14: Tyr4 C-C_γ Konformationsvorschlag. Konformation des dritten C-C_γ Diederwinkels. Gezeigt sind außerdem die sterischen Konflikte, die bei der Benutzung als Startstruktur entstanden wären. Die roten Plättchen zeigen dabei starke, entsprechend ihrer Größe, paarweise Überlappungen von atomaren van-der-Waals Radien an.

Die dritte Konformation wird aber bei der Modellierung verworfen, da sie mit zu vielen sterischen Konflikten einhergeht (Abbildung 3.14). Diese treten vermutlich nicht innerhalb einer Simulation auf, da das System bzw. entsprechende Seitenketten ausweichen können. Dies ist ein weiteres Beispiel für die Entdeckung von kryptischen Bindungsstellen während Molekulardynamiksimulationen [178]. Bei der Orientierung des Tyr4 Phenolrings zeigt dieser zwischen die RT-Schleife und dem β d-Strang (Zur Beschreibung der Nomenklatur von SH3 Domänen, siehe Abbildung 1.4). Diese Region ist bekannt dafür, variabel und flexibel zu sein [38, 179]. Daher wird sie oft in wissenschaftlichen Studien untersucht [180, 181, 182, 183, Beispiele für Hck-SH3]. Außerdem können in diesem Bereich die Aminosäuren E35, H38 und H39 der RT-Schleife nicht durch chemische Verschiebungen im Experiment nachgewiesen werden [37]. Das erhärtet den Verdacht dieser zusätzlichen Bindungsmöglichkeit.

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

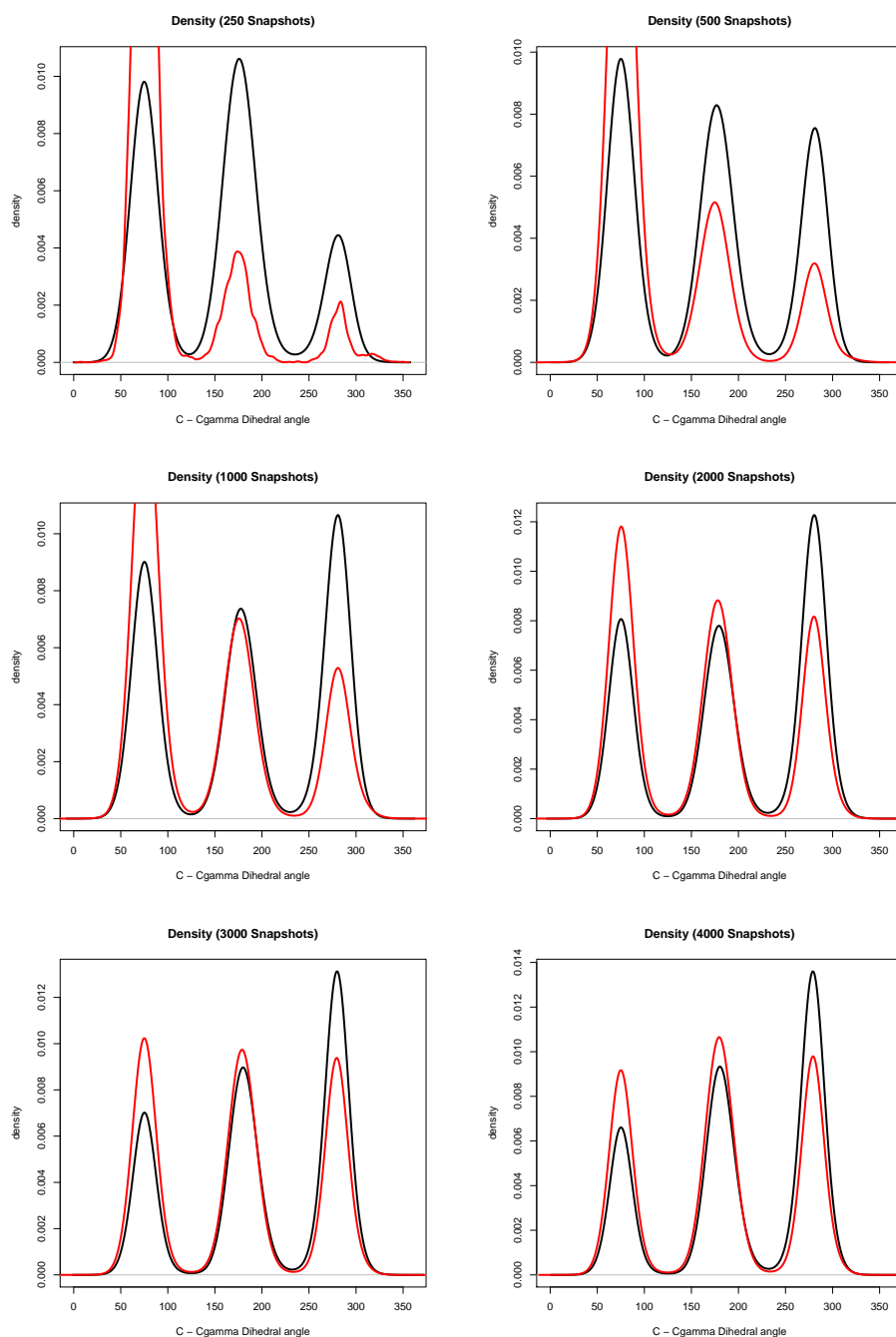


Abbildung 3.15: C-C γ Dichteverteilungen des Tyr4 C-C γ Diederwinkels. Die Dichteverteilungen sind für 2, 5 ns, 10 ns, 20 ns, 30 ns und 40 ns gezeigt, was 250, 500, 1000, 2000, 3000, 4000 Zeitschritten, hier Snapshots, entspricht. Die Verteilung der Referenzstruktur ist schwarz dargestellt, die der Modellstruktur rot.

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

Auftragungen von Dichteverteilungen zu verschiedenen Zeitpunkten der Simulation zeigen die Entwicklung der drei Konformationen (Abbildung 3.15). Man erkennt deutlich die Entwicklung der Verteilungen in Abhängigkeit von ihrer Startstruktur. Nach 2,5 ns Simulationszeit, in Abbildung 3.15 (oben links) als 250 Snapshots bezeichnet, ist die Dichteverteilung der Referenzstruktur, schwarze Kurve, bei $150 - 200^\circ$ am höchsten. Dass die Dichteverteilung bei einem Winkel von $50 - 100^\circ$ nahezu gleich hoch ist zeigt, dass diese Konformation leicht und ohne die Anpassung anderer Elemente in der Struktur zu erreichen ist. Die Modellstruktur startet hingegen bei $50 - 100^\circ$. Die rote Kurve zeigt, dass die anderen Konformationen von hier aus nicht einfach zu erreichen sind, da die Dichteverteilungen über den anderen Winkeln nur klein sind. In den Graphen der anderen Dichteverteilungen erkennt man deutlich, dass Verteilungen über der Zeit stark schwanken. Die meisten Änderungen finden in den ersten 10 ns bzw. 1000 Snapshots statt. Die Änderungen in den letzten 10 ns bzw. zwischen 3000 und 4000 Snapshots bleiben hingegen überschaubar. Dass in beiden Fällen die Konformation bei $300 - 350^\circ$ erst nach 10 ns (schwarz) bzw. 20 ns (rot) oft erreicht wird, mag zum einen an den oben erwähnten Strukturänderungen liegen, ist aber auch durch die Tatsache bedingt, dass dieser Winkel nur über den Winkel $150 - 200^\circ$ erreichbar ist. Nur vier der Übergänge erfolgen direkt aus dem Winkel $50 - 100^\circ$, alle anderen kommen aus dem Winkel $150 - 200^\circ$ (Daten nicht gezeigt).

Eine andere Herangehensweise, den Tyr4 Diederwinkel zu untersuchen, ist in Abbildung 3.16 gezeigt. Auch hier erkennt man von welchem Winkel gestartet wurde, und im Fall der Referenzstruktur, einen zeitnahen Wechsel in die beiden anderen Konformationen. Sowohl im oberen, als auch im unteren Graphen lässt sich erahnen, dass sich die Modellkonformation bei etwa 20 % einpendeln wird. Die Verteilung der anderen beiden Winkel ist im oberen und unteren Graphen vertauscht. Ein Wechsel der blauen und roten Kurve lässt sich zu einem späteren Zeitpunkt nur vermuten. Gut zu erkennen ist hingegen bei der Modellstruktur (obere Abbildung 3.16), dass die kryptische Konformation von über 240° nur langsam erreicht wird. Vermutlich sind die Kurven aus diesem Grund auch hier vertauscht, da dieser Diederwinkel nur aus einer Richtung erreichbar ist. Deutlich wird aus dem Verlauf beider Graphen, dass eine gleiche Verteilung der Konformationen innerhalb akzeptabler Simulationszeiten nicht möglich ist. Aus diesem Grund ist

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

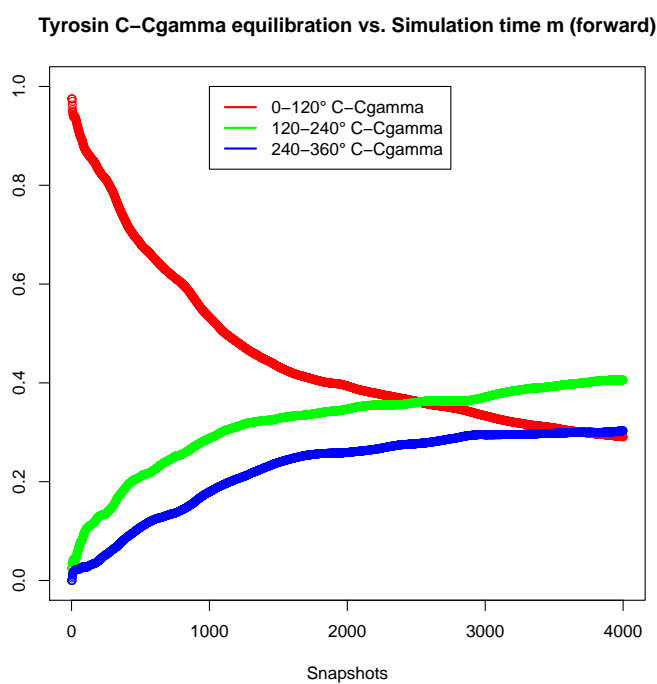
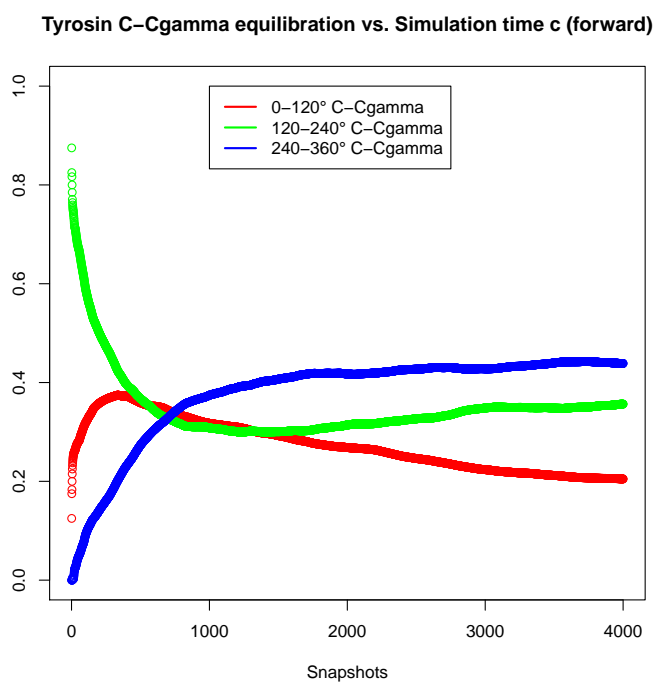
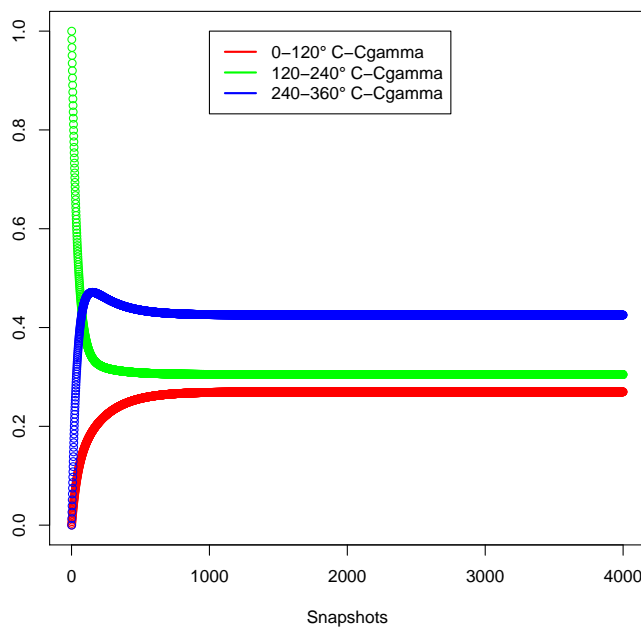


Abbildung 3.16: Entwicklung des Tyr4 C-C γ Diederwinkels. Einteilung aller eingenommenen Tyr4 C-C γ Diederwinkel in drei Konformationen. Zusammenfassend für alle Simulationen der Referenzstruktur (oben) und der Modellstruktur (unten).

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

Tyrosin C–Cgamma equilibration from 2000 Snapshots for c (predicted)



Tyrosin C–Cgamma equilibration from 2000 Snapshots for m (predicted)

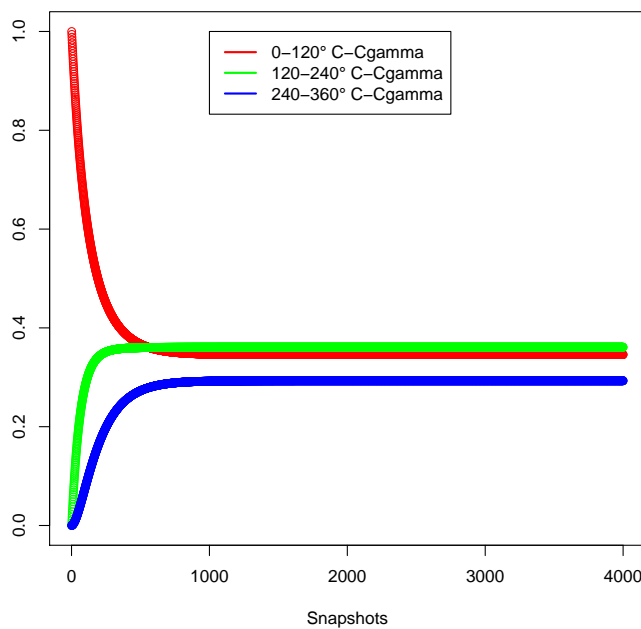


Abbildung 3.17: Vorhersage des Tyr4 C–C_γ Diederwinkels. Einteilung aller eingenommenen Tyr4 C–C_γ Diederwinkel in drei Konformationen. Zusammenfassend für alle Simulationen der Referenzstruktur (oben) und der Modellstruktur (unten).

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

dies erst recht nicht gleichzeitig für alle Konformationen im System zu erwarten. Da die Seitenkette niemals zwei Konformationen gleichzeitig einnehmen kann und die Kurvenverläufe in Abbildung 3.16 an lineare Populationsdynamiken erinnert, wird im Folgenden untersucht, ob die Winkelkonformationen nicht durch eine solche Methode berechnet werden können. Abbildung 3.17 zeigt die berechneten Kurvenverläufe. Bei einem Vergleich mit Abbildung 3.16 zeigt sich zwar für beide Startkonformationen der gleiche Trend. Aber in beiden Vorhersagen stellt sich ein verfrühtes Gleichgewicht innerhalb der ersten 1000 Zeitschritte ein. Außerdem erfolgt in der Vorhersage zur Modellstruktur keine Kreuzung der roten und blauen Kurve, wie es in den letzten 500 Zeitschritten der Simulation geschehen ist. Beide Beobachtungen legen nahe, dass die Entwicklung der Winkelkonformationen keiner linearen Abhängigkeit gehorcht. Dies ist aber Voraussetzung für die Anwendung der Populationsdynamik. Möglicher Grund dafür ist, dass nicht jeder Winkel beliebig oft eingenommen werden kann, wie in der obigen Diskussion bereits beschrieben.

MMGBSA als Konformationsanalyse Eine schnelle und grobe Methode, im Gegensatz zu anderen Methoden wie der thermodynamischen Integration, um Unterschiede in Bindungsaffinitäten zu analysieren, ist das MMGBSA Verfahren (Kapitel 2.6.2). Daher wird diese Methode im Folgenden verwendet, um mögliche Abhängigkeiten von Startstrukturen und Bindungsaffinitäten zu untersuchen. Tabelle 3.10 zeigt die Ergebnisse aller Simulationen der Referenz- und Modellstruktur. Die Tabelle enthält eine obere und eine untere Hälfte. In der oberen Hälfte wird die Menge der analysierten Zeitschritte variiert. Hier bedeutet 10 *ps* dass die

Tabelle 3.10: MMGBSA Ergebnisse in Abhängigkeit der Startstruktur.

	Referenzstruktur	Modellstruktur
10 <i>ps</i>	-61,47	-60,95
100 <i>ps</i>	-61,50	-60,43
1000 <i>ps</i>	-61,45	-60,08
0 – 120°	-60,43	-62,53
120 – 240°	-59,83	-58,74
240 – 360°	-60,13	-61,86

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

Zeitschritte der Trajektorien alle 10 ps ausgewertet werden, entsprechendes gilt für 100 ps und 1000 ps . Es fällt auf, dass die Referenzwerte dabei um einen kleineren Wert schwanken als die Ergebnisse der Modellstruktur. Bei einem Vergleich mit Abbildung 3.16 erscheint dies logisch, da sich hier die Kurven erst vor Ende der Simulationen kreuzen. Dies geht einher mit der Menge der eingenommenen Konformationen, wobei die Konformationen selbst aber auch unterschiedliche Affinitäten besitzen können. Um diese beiden Fälle zu untersuchen, werden im unteren Teil der Tabelle 3.10 die 10 ps Datensätze nach den drei Rotameren aufgeteilt und analysiert. Bei beiden Datensätzen zeigt sich der gleiche Trend der Affinitäten, so dass eine winkelabhängige Affinität wahrscheinlich ist. Es zeigt sich aber wiederum dass die Werte bei der Modellstruktur weiter auseinander liegen. Da an dieser Stelle das Projekt aus zeitlichen Gründen nicht tiefer bearbeitet werden kann, bleibt die Vermutung im Raum stehen, dass die Unterschiede von der Größe der Datensätze abhängen. Eine andere Untersuchung könnte bestätigen, dass die anderen Teile der Struktur, bei gleichem Diederwinkel, in unterschiedlichen Teilen der Trajektorien gleich sind. Dies ist bis dato lediglich angenommen worden.

SAM Analyse Ausgehend von verschiedenen Rotameren zeigen die vorangegangenen Ergebnisse, dass keine gleichen Verhältnisse der möglichen Konformationen dargestellt werden können. Darüber hinaus sagen die bisherigen Untersuchungen nichts über das weitere Verhalten des Systems aus. Bei genauerer Betrachtung ist es für den praktischen Nutzen von Molekulardynamiksimulationen auch nicht notwendig, das Verhalten anderer Simulationen zu imitieren. Wichtig ist es lediglich, ein von der Startstruktur unabhängiges repräsentatives Verhalten des zu untersuchenden Systems aufzuzeigen. Mit den bisherigen Methoden ist es nicht möglich das gesamte System auf sein repräsentatives Verhalten zu untersuchen. Dies wäre aber mit einer Signifikanzanalyse wie SAM (Kapitel 2.6.5) möglich.

Die Methode wird nun verwendet um zu klären, ob in den Simulationen der zwei Startkonformationen für diskrete Zeitfenster signifikante Unterschiede zu finden sind. Mit diesem Verfahren ist es möglich, über den Tellerrand der einzelnen Unterschiede, hier Tyr4 C-C γ Diederwinkel, auf alle Unterschiede im gesamten System zu betrachten. Mit dieser Art der Analyse ist es ohne weiteres möglich, einen,

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

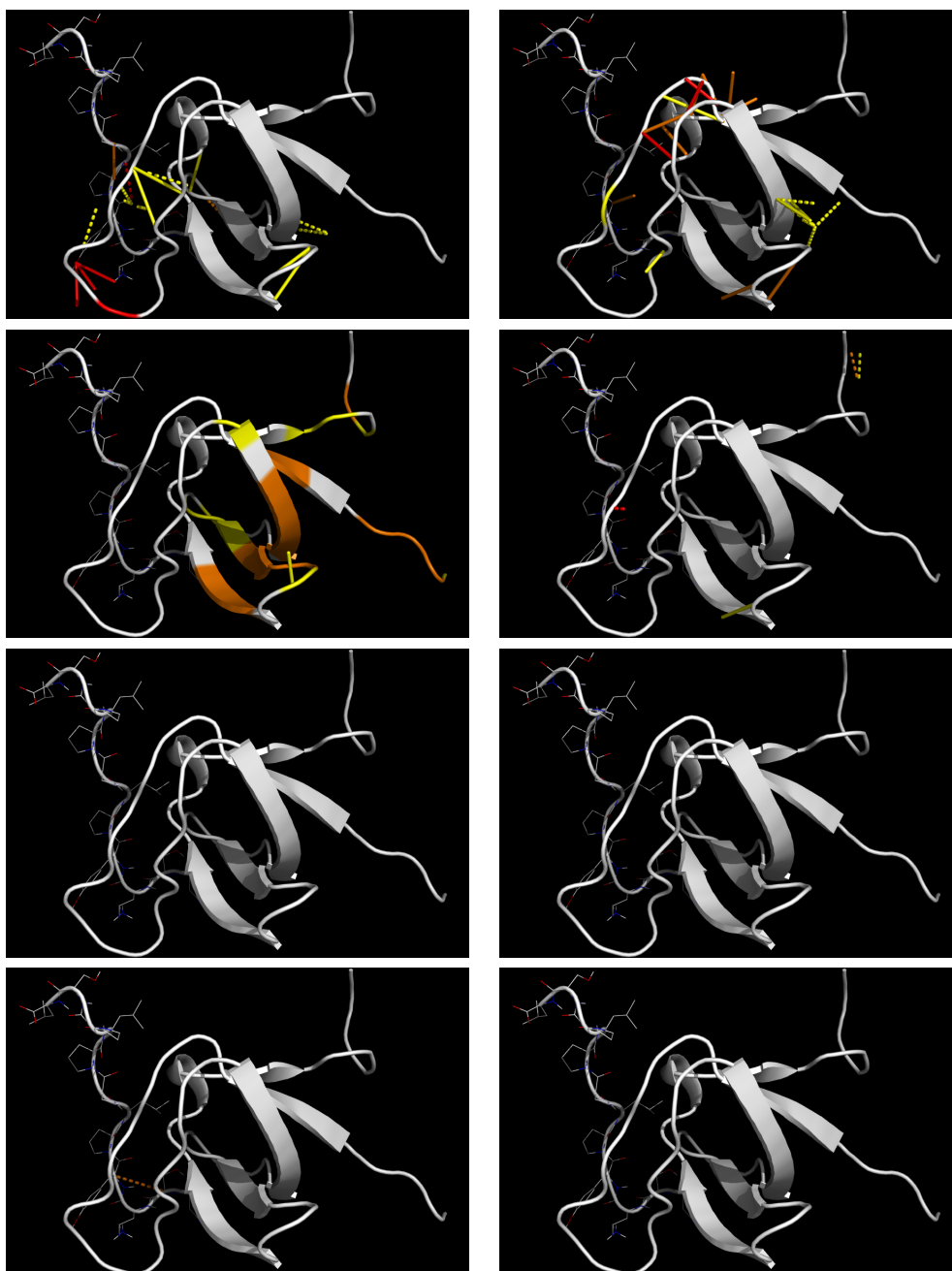


Abbildung 3.18: Simulationsvergleich der Referenz- und Modellstruktur mit SAM in 5ns Zeitfenster. Die einzelnen Strukturen werden von oben links, 0 – 5ns, nach unten rechts, 35 – 40ns, gelesen. Neben der Sekundärstruktur des Peptid-Protein-Komplexes werden die Peptidseitenketten als Linien mit Tyr4 in der unteren linken Ecke des jeweiligen Komplexes dargestellt. Die Farben und der Darstellungscode entsprechen dem der Abbildung 3.19.

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

wie hier verwendeten, Simulationsansatz mit mehreren unabhängigen Trajektorien zu behandeln. Dies ist in letzter Zeit mehr und mehr zur Herausforderung geworden ist. Abbildung 3.18 zeigt den Vergleich von Referenz- und Modellstruktur, eingeteilt in 5 ns Zeitfenster. Das bedeutet, dass die oberste linke Struktur alle Unterschiede in den Trajektorien von $0 - 5\text{ ns}$ zeigt, die oberste rechte Struktur von $5 - 10\text{ ns}$ usw., bis zur untersten rechten Struktur, die die letzten 5 ns von $35 - 40\text{ ns}$ zeigt. Die Relevanz der signifikanten Unterschiede ist dabei mit einer Farbkennzeichnung - von rot, für hoch signifikant (1% FDR), über orange, nach gelb, für gering signifikant (5% FDR) - gelöst worden. Eine gefärbte Sekundärstruktur zeigt signifikante Unterschiede in den RMSF Werten. Signifikante Distanzunterschiede (siehe Kapitel 3.3.2) sind durch Linien gekennzeichnet. Unterbrochene Linien beschreiben Unterschiede in der Varianz der Werte, durchgezogene Linien stellen die Durchschnittswerte dar. Es zeigt sich, dass die Unterschiede von einem lokalen hoch signifikanten Ort (mit den drei roten Linien in der obersten linken Figur, in dem sich das Tyr4 befindet) über die RT-Schleife im $5 - 10\text{ ns}$ Zeitfenster zur Faltblattstruktur im $10 - 15\text{ ns}$ Zeitfenster wandert und dann mit nur noch minimalen Unterschieden im $15 - 20\text{ ns}$ Zeitfenster aus der Struktur fließt. Nach 20 ns Simulationszeit gibt es keine Unterschiede mehr zwischen den beiden Sätzen von Trajektorien. Das heißt, beide Ansätze zeigen keine statistisch signifikanten Unterschiede mehr. Damit ist es nicht mehr möglich, die Ausgangsstruktur zu identifizieren. Konfigurationen nach diesem Zeitpunkt können sowohl von der Referenz- als auch von der Modellstruktur stammen. Damit ist diese Methode ideal, um Simulationsansätze mit mehreren Trajektorien zu analysieren. Praktisch gesehen schließt sie somit Interpretationen auf Grundlage von Modellierungsfehlern aus. Da sich in den MMGBSA Rechnungen (Tabelle 3.10) die Affinitäten eines bestimmten Winkels zwischen den Referenz- und Modellstruktursimulationen unterscheiden, wird dieser Zusammenhang im Folgenden strukturell untersucht und in Abbildung 3.19 qualitativ zusammengefasst. Hier zeigt sich das gleiche Szenario wie in Abbildung 3.18, in dem Unterschiede nur in der ersten Hälfte der Trajektorie zu finden sind. Eine Ausnahme bilden lediglich einige geringe, gelb gekennzeichnete RMSF Unterschiede der $0 - 120^\circ$ Strukturen. Dies unterstützt die Interpretation der MMGBSA Ergebnisse. Dort würde eine Unterscheidung in zumindest zwei Zeitfenstern die Ergebnisse wahrscheinlich angleichen.

3.3 Vergleich von Molekulardynamiksimulationen mit unterschiedlichen Startstrukturen

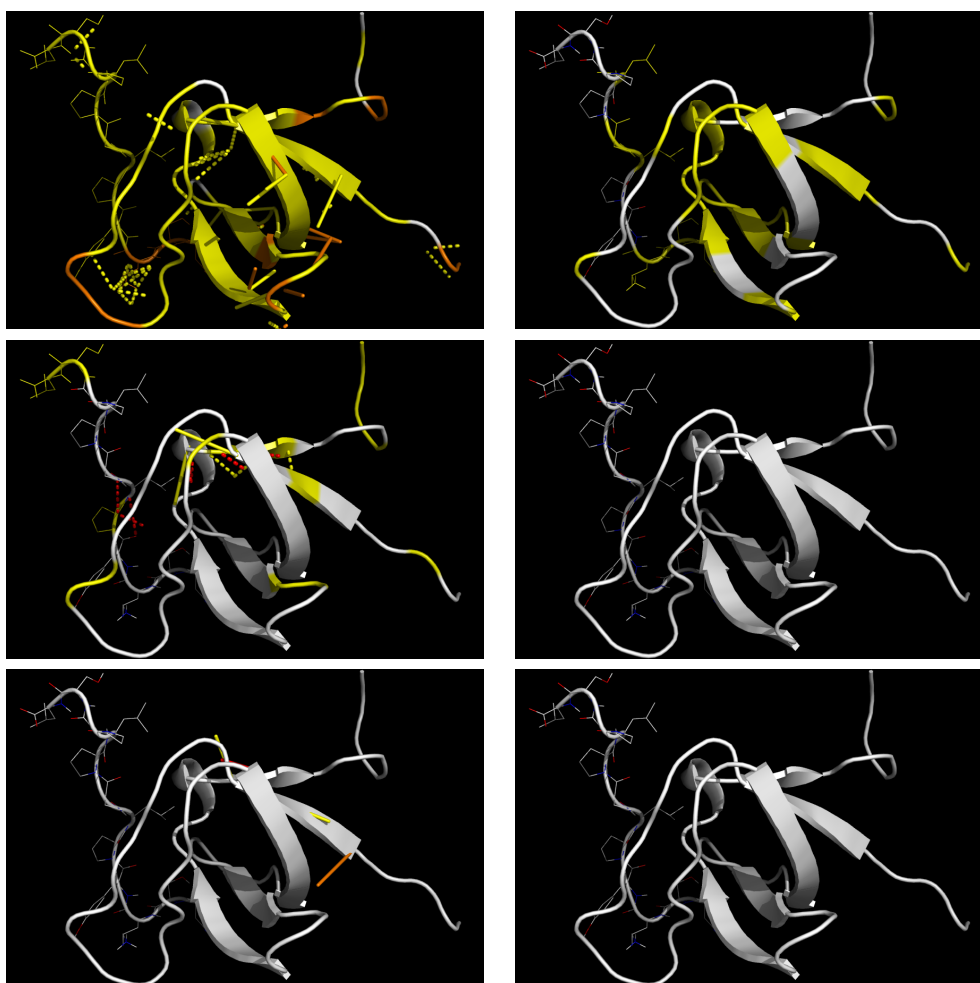


Abbildung 3.19: Simulationsvergleich der Referenz- und Modellstruktur mit SAM im 5 ns Zeitfenster und in Abhängigkeit des Tyr4 C-C γ Diederwinkels. Die oberste Reihe der Figuren vergleicht die Strukturen mit einem Tyr4 C-C γ Diederwinkel von 0 – 120°. In der mittleren Reihe sind die Strukturen mit einem Tyr4 C-C γ Diederwinkel von 120 – 240° und in der untersten Reihe mit einem Tyr4 C-C γ Diederwinkel von 240 – 360° zu finden. Alle linken Figuren zeigen dabei Strukturen in der ersten Hälfte (von 0 – 20 ns) der Simulationszeit und alle rechten Figuren in der zweiten Hälfte (von 20 – 40 ns) der Simulationszeit. Die Farben rot, orange und gelb kennzeichnen den Grad der signifikanten Unterschiede hoch, mittel und niedrig. Unterschiede in der RMSF werden als Farben in der Sekundärstruktur angezeigt. Unterschiede in den C α -, van-der-Waals-, NO- oder HO-Distanzen werden als Linien dargestellt. Eine gestrichelte Linie beschreibt Unterschiede in den Varianzwerten, wobei durchgezogene Linien Unterschiede in den Mittelwerten beschreiben.

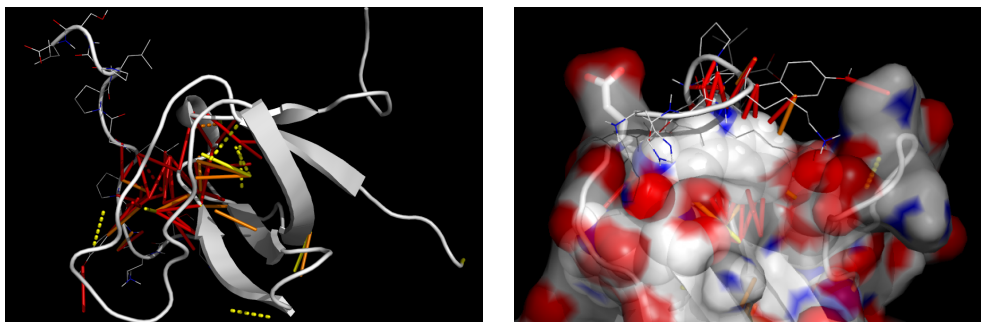


Abbildung 3.20: Vergleich der Winkelkonformationen 120 – 240° und 240 – 360°. OBEN: Genutzt werden ausschließlich die Simulationen mit der Referenzstruktur als Startkonformation für die Simulationszeit von 5 – 40 ns. Die Farbe und der Darstellungscode entsprechen dem der Abbildung 3.19. UNTEN: Nahaufnahme der signifikanten Unterschiede in der Mitte der Kontaktfläche von Peptid und SH3 Domäne. Die in Kugeln dargestellte Aminosäure ist in allen SH3 Domänen konserviert und ist in die meisten signifikanten Unterschiede involviert.

Abschließend wird SAM verwendet, um die Strukturen mit der kryptische Konformation zu untersuchen. Dazu werden Strukturen aus den Referenzsimulationen mit einem Winkel von $240 - 360^\circ$ gegen Strukturen mit einem Winkel von $120 - 240^\circ$ verglichen. Um uninteressante Unterschiede zu vermeiden, werden die Trajektorien erst ab 5 ns Simulationszeit ausgewertet. Es zeigt sich in der linken Abbildung 3.20, dass die größten Unterschiede in der Mitte der Kontaktfläche zwischen Peptid und SH3 Domäne zu finden sind. Die Seitenkette des Tyr4 ist dabei in der linken unteren Ecke der Referenzstruktur zu finden. Bei einer Rotation in die kryptische Konformation zeigen sie an dieser Stelle die Mehrzahl der roten signifikanten Unterschiede. Die rechte Struktur (Abbildung 3.20) zeigt, dass die meisten dieser Unterschiede die Konformation einer anderen, als Kugeln dargestellten, Aminosäuren beeinflusst. Die Aminosäure ist das prominente und konservierte Tryptophan, welches in jeder SH3 Domäne gefunden wird [38]. Eine bewegliche W59 Seitenkette wurde schon durch Fernandez-Ballester et al. [36] beschrieben, jedoch in einem anderen Zusammenhang (siehe Kapitel 1.2.3). Was die hier gefundenen Ergebnisse für jene Ergebnisse bedeuten, ist eine interessante Fragestellung für zukünftige Untersuchungen.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

3.4.1 Einführung

Viele Anstrengungen wurden auf die Gestaltung und Identifizierung von hochaffinen Liganden für SH3 Domänen verwendet. Hierbei wurde mit den unterschiedlichsten Methoden gearbeitet [31]. Beispielsweise haben Tran et al. [184] durch das Screenen einer Phagen-Display-Peptidbibliothek mehrere künstliche Peptidliganden für die Lck (Lymphozytenspezifische Kinase)-SH3 Domäne identifiziert. In deren Arbeit enthielten die meisten der ausgewählten Peptide das Klasse I Motiv (Kapitel 1.2.3), wobei keines von ihnen eine positiv geladene Aminosäure an P₋₃ enthielt. Die Bindungsaffinitäten des dominierenden Peptids, genannt PD1 (HSKYPLPPLPSL), zeigten Affinitäten zu den SH3 Domänen von Lck, Hck, Fyn, Src und PI3K, die zwischen 0,2 μM und 120 μM variierten (Tabelle 3.11). Dabei wurde der stärkste Komplex von PD1:Hck-SH3 und der schwächste Komplex von PD1:PI3K-SH3 gebildet [37, 185]. Für eine bessere Unterscheidung zwischen SH3 Domäne und Peptid wird im Folgenden der Ein-Buchstaben-Code für die Aminosäuren der SH3 Domäne und für die Aminosäuren des Peptids der Drei-Buchstaben-Code verwendet. Schmidt et al. [37] haben eine NMR Struktur des stärksten Komplexes (Tabelle 3.11), PD1:Hck-SH3, mit Tyr4 in der P₋₃ Ankerposition (Kapitel 1.2.3) experimentell zeigen können. Diese Struktur zeigte bis dato unbekannte Interaktionen, da PD1 durch einen Knick im Peptidrückgrat Lys3, Peptidposition P₋₄, in die Kompassstasche von Hck-SH3 (Abbildung 3.21) bringt.

Tabelle 3.11: Überblick über die in dieser Arbeit verwendeten Peptid-SH3 Domänen. Als Referenzstrukturen ist eine PDB erhältlich. Da nur die Peptid-SH3 Domänen Komplexe mit der stärkeren Affinität als PDB zur Verfügung stehen, werden die Modellstrukturen durch eine Punktmutation im Peptid der jeweiligen SH3 Domänen PDB gebildet.

	Hck SH3 Domäne [37]		PI3K SH3 Domäne [185]	
	K _D Wert (μM)	Struktur	K _D Wert (μM)	Struktur
PD1	0,2	Referenz	120	Modell
PD1R	0,5	Modell	40	Referenz

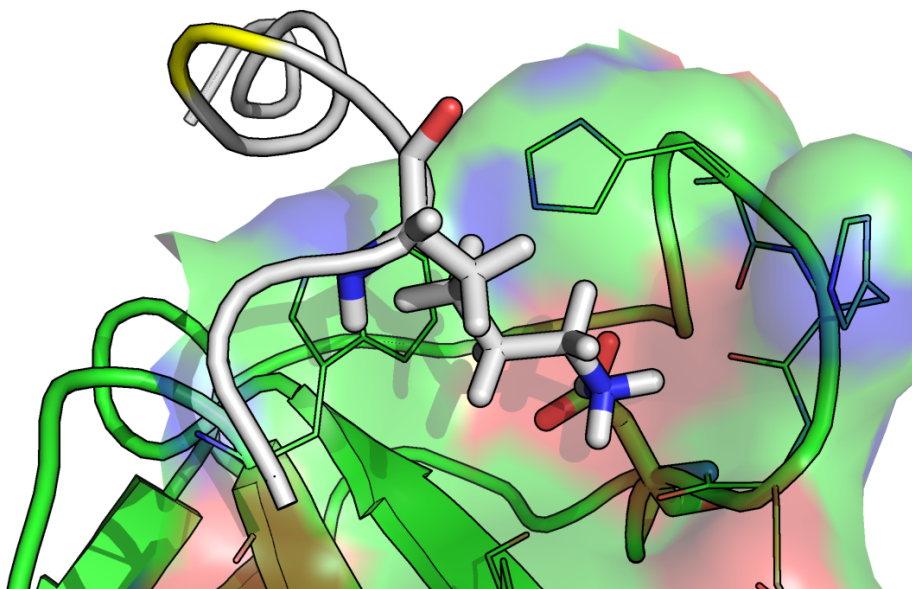


Abbildung 3.21: Knick und Kompassstasche der PDB Struktur PD1:Hck-SH3. Das Peptid ist grau dargestellt, das Protein grün. Im Hintergrund ist die PPII Peptidstruktur zu erkennen. Im Vordergrund ist der Knick im Peptidrückgrat in gelb dargestellt. Durch den Knick ist es möglich, Lys3 (Stabmodell) in die Kompassstasche (transparente Oberfläche mit entsprechenden Aminosäuren als Linien) zu positionieren und mit D21 (Stabmodell) am Boden der Tasche zu interagieren.

An dieser Stelle muss darauf hingewiesen werden, dass Schmidt et al. [37] keine intermolekularen NOE Signale zwischen den Seitenkettenprotonen von Lys3 und Protonen der Hck-SH3 Domäne sehen konnte. Diese werden normalerweise dazu genutzt, um den Abstand zwischen den Atomen bestimmen zu können. Dennoch deuten weitere Berechnungen und Mutationsexperimente darauf hin, dass die Seitenkette von Lys3 in die Kompassstasche zeigt.

Die PD1 Variante PD1R (HSKRPLPPLPSL) trägt den basischen Anker als Punktmutation Tyr4Arg. Die Bindungskonstanten von PD1R zu den SH3 Domänen von Lck, Hck, Src, Fyn und PI3K Kinasen variieren von $0,5 \mu\text{M}$ bis $40 \mu\text{M}$ [37, 185]. Damit variieren die Affinitäten von PD1R einen kleineren Bereich als die von PD1 (Tabelle 3.11), wobei der stärkste Komplex wiederum von Hck-SH3 und der schwächste von PI3K-SH3 gebildet wird. Batra-Safferling et al. [185] haben eine

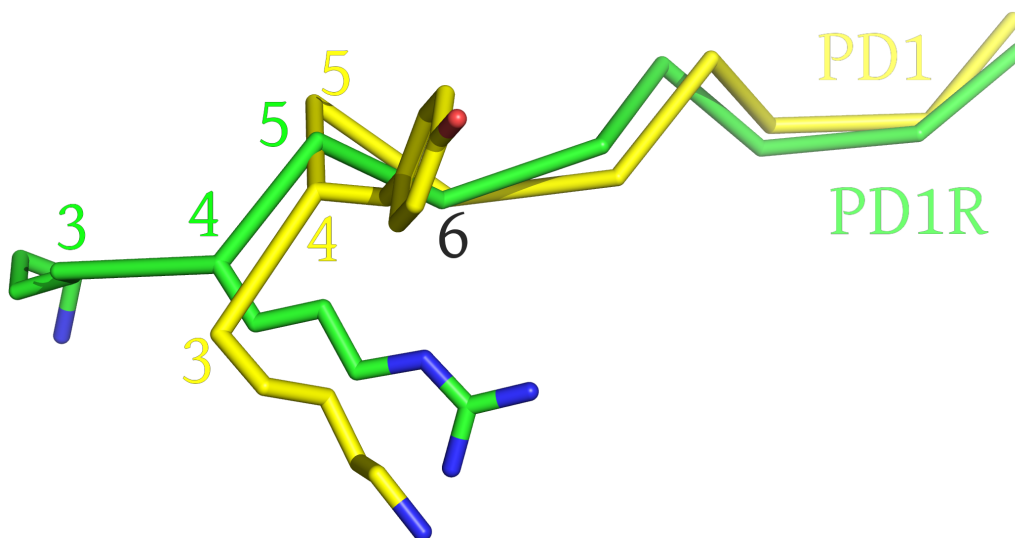


Abbildung 3.22: Strukturelle Überlagerung des Peptidrückgrats von PD1 (gelb) und PD1R (grün). Die Überlagerung lässt deutlich einen Knick an Position 5 von PD1 erkennen. Dies ermöglicht es Lys3 (Position 3; P₋₄) als Anker zu agieren. Dies ist normalerweise, wie auch im Fall von PD1R, Aminosäuren an P₋₃ (Position 4 in PD1R) vorbehalten. In PD1R zeigt Lys3 (Position 3) keine wichtigen Interaktionen mit PI3K-SH3 und ist somit in die entgegengesetzten Richtung orientiert. Das entspricht einer Ausrichtung ins Lösemittel. Ab Position 6 verlaufen beide Peptide nahezu parallel.

Kristallstruktur von PD1R im Komplex mit der PI3K-SH3 Domäne publiziert. In dieser Struktur ist Arg4 (Tabelle 1.2) in Richtung der Kompassstasche orientiert. Außerdem zeigt im Gegensatz zu PD1 im PD1:Hck-SH3 Komplex (Abbildung 3.22) der Ligand keinen Knick in seinem Peptidrückgrat. Die Arg4 Konfiguration steht im Einklang mit dem allgemeinen Bindungsmodus der Klasse I Peptide mit SH3 Domänen (Kapitel 1.2.3). Ferner verglich die Arbeit von Batra-Safferling den Komplex PD1R:PI3K-SH3 mit PD1:Hck-SH3 und fand die hydrophoben Teile (Aminosäuren 5-12) beider Peptide in einer ähnlichen Konfiguration. Ausschließlich die Aminosäuren 1-4 zeigten Unterschiede in ihren Konfiguration, welche den Knick des Rückgrats und damit die verschiedenen Aminosäuren in den Kompassstaschen mit einschließen.

Der Versuch, PD1:PI3K-SH3 strukturell aufzuklären, resultiert in einem PI3K-SH3 Dimer mit jeweils R14 (Nummerierung der Aminosäuren aus [38]) übernom-

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

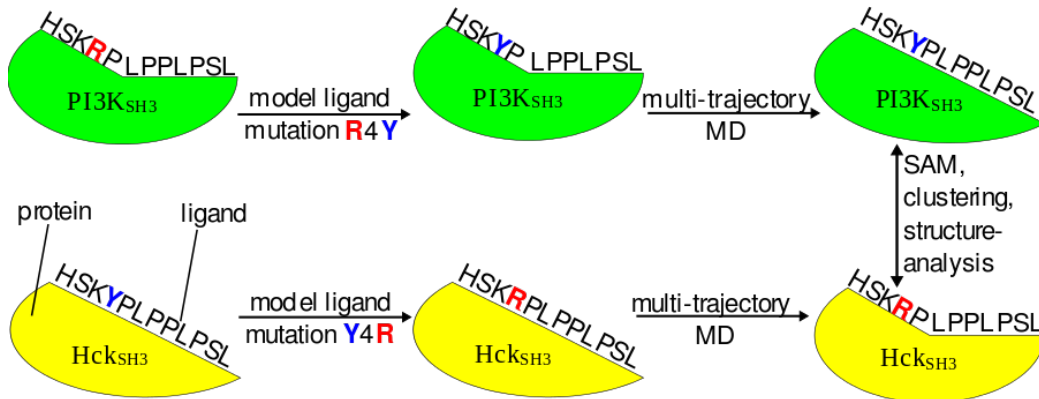


Abbildung 3.23: Workflow in diesem Kapitel.

men) in der Kompassstasche. Dieses Dimer beschreibt eine $\pi - \pi$ Wechselwirkung, die oft in Kristallstrukturen gefunden wird [185]. Ebenso zeigten Schmidt et al. [37] den Hck-SH3 Komplex mit der niedrigsten Affinität von Tran et al. [184, 185] nicht. Sie vermuteten aber, dass die Punktmutation Tyr4Arg, und damit das Peptid PD1R, teilweise in der Lage sei, die Wechselwirkungen von Tyr4 zu ersetzen [37]. Basierend auf den drei Jahre später gelösten Komplex PD1R:PI3K-SH3 sind die Annahmen von Schmidt et al. [37] zweifelhaft. Das Wissen beider Komplexe mit niedriger Affinität (PD1:PI3K-SH3, PD1R:Hck-SH3) basiert somit ausschließlich auf Annahmen. Daher werden im Folgenden diese Vermutungen, mittels eines Workflows aus: *in silico* Punktmutationen, Molekulardynamiksimulationen, Signifikanzanalysen und Clusteranalysen untersucht und aufgeklärt (Abbildung 3.23).

Unter der Annahme, dass die Peptide in beiden unbekannten Strukturen auch als Klasse I Liganden binden werden, macht es in der Praxis Sinn, nur Punktmutationen in die PDB Strukturen bzw. Referenzen (PD1R:PI3K-SH3, PD1:Hck-SH3) einzuführen und damit die Modelle für PD1:PI3K-SH3 und PD1R:Hck-SH3 zu generieren. Deshalb startet die Generierung der Modelle auch nicht mit einer neuen Modellierung wie in Kapitel 1.2.1, oder einer Generierung des Protein durch Faltungsexperimente [186, 187] zu erzeugen, um dann das Peptide mit ungelenkten Simulationen [188, 189, 190] an seine Bindungsstelle zu koordinieren. Dies hätte einen ungerechtfertigt großen Rechenaufwand bedeutet. Trotzdem kann die Modellierung von Punktmutationen an sich schon einen hohen Rechenaufwand

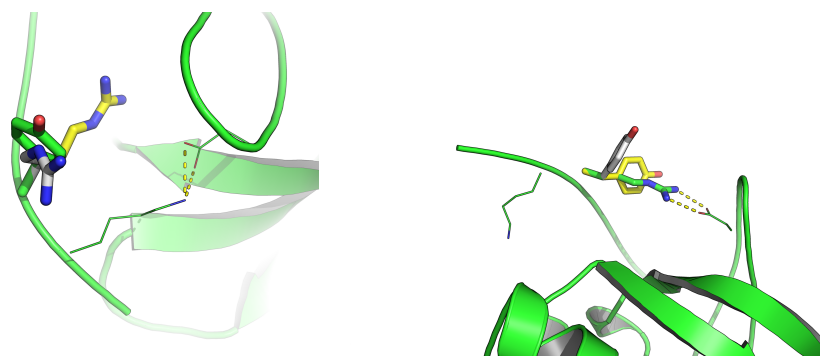


Abbildung 3.24: Startkonformationen von Arg4 in PD1R:Hck-SH3 (links) und von Tyr4 in PD1:PI3K-SH3 (rechts). In beiden Figuren ist die Salzbrücke am Boden der Kompass tasche eingezeichnet.

bedeuten, da sie in der Umlagerung von Seitenketten in der Region der Mutation resultieren kann (Kapitel 3.3). Auch wenn bisher verschiedenste Methoden zur Modellierung verwendet wurden, wurden auch immer wieder Molekulardynamiksimulationen zur Verfeinerungen eingesetzt [191, 112, 192, 193]. Meistens wurde in diesen Fällen, wie in Kapitel 3.3, ein Ansatz mehrerer unabhängiger Simulationen verwendet, die darüber hinaus Ansätze mit nur einer Simulation übertrafen, auch wenn sie in der Summe eine kürzere Simulationszeit hatten [114, 112, 194, 67]. Leider war dabei nicht immer nachvollziehbar, warum bestimmte Strukturen als repräsentativ ausgewählt wurden. Häufig wurde das letzte Bild einer Simulation als repräsentativ angenommen [112, 113, 114]. Dies erscheint eher willkürlich. Solch ein Ansatz ist unzureichend, da bekannt ist, dass Simulationen nie vollständig äquilibrieren [115, 116, 117, 113]. Darüber hinaus steht es im Konflikt mit der Ergodenhypothese, die ein repräsentatives Ensemble annimmt.

In vorherigem Kapitel wurde gezeigt, dass man mit einer statistischen Methode wie der SAM Methode von Tusher et al. [118] diese Probleme vermeiden und statistisch repräsentative Abschnitte in Trajektorien gefunden werden können. Dazu wird SAM auch hier verwendet. Mit dem erlangten Wissens wird diesmal ein Clusteralgorithmus (Kapitel 2.6.4) gefüttert, welcher die Datensätze für eine strukturelle Analyse weiter zusammenfasst (Abbildung 3.23). Die vorgeschalteten Simulationen starten, wie in Kapitel 3.3, mit verschiedenen Rotameren (Ab-

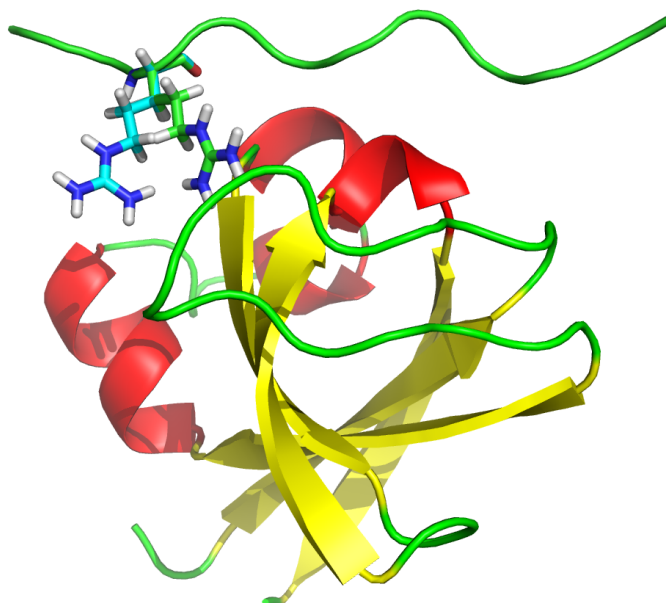


Abbildung 3.25: Unterschied in den Startstrukturen von PD1R:PI3K-SH3. Der einzige Unterschied besteht in der rotierten Seitenkette von Arg3. Dabei führte die Rotation dazu, dass die Seitenkette nicht mehr in der Kompass tasche koordiniert ist.

bildungen 3.24), hier aber mit Rotameren der eingeführten Punktmutationen. Um noch mehr Wissen über SAM zu erlangen, wird neben dem aus dem vorherigen Kapitel 3.3 untersuchten Datensatz ein weiterer Datensatz mit SAM untersucht. Es ist die zweite Referenzstruktur PD1R:PI3K-SH3 (Tabelle 3.11 und Abbildung 3.25). Sie wird zur Modellierung von PD1:PI3K-SH3 eingesetzt, wobei PD1R:Hck-SH3 aus PD1:Hck-SH3 modelliert wird (Abbildung 3.23). Dabei werden Variablen wie die gebrauchte Menge von Trajektorien, Permutationsgröße und FDR untersucht. Diese Erkenntnisse fließen anschließend in die Untersuchung der Modellstrukturen mit ein.

3.4.2 Material und Methoden

Strukturvorbereitung Alle Modellierungsschritte werden in der gleichen Weise wie in Kapitel 3.3.2 beschrieben durchgeführt. Die verwendeten PDB Strukturen sind [185, PDB ID: 3i5r] und [37, PDB ID: 2oi3]. Im Peptid von 3i5r sind die Aminosäuren 1, 2 und 12 zusätzlich modelliert, da sich in Testsimulationen

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

gezeigt hat (Daten nicht gezeigt), dass die mittleren Aminosäuren der Peptidkette niedrigere RMSF Werte aufweisen. Dies ist die Basis für stabile Interaktionen.

Simulationsprotokoll Alle Simulationen der Referenzstruktur PD1R:PI3K-SH3 werden mit der Amber11 [85] Software durchgeführt. Durch ein Update von Amber11 auf Amber12 werden die Modelle, PD1R:Hck-SH3 und PD1:PI3K-SH3, mit Amber12 [195] simuliert. Sander wird dabei für die Energieminimierung und PMEMD für die Simulation verwendet. Alle Strukturen werden mit leap von AmberTools 1.5 [85] vorbereitet. Dabei wird das Kraftfeld Amber99SBildn [172] verwendet. Die Protonierungszustände aller PI3K-Komplexe werden mit dem PROPKA [196] Webserver berechnet. Die Protonierungszustände der Hck-Komplexe werden der NMR Struktur entnommen. Ungeladene Histidin Seitenketten werden mit HID, entsprechende Amber Nomenklatur, in der PDB umgeschrieben. Der Komplex wird in eine oktaedrische Box mit TIP3P [173] Wasser getaucht, wobei ein Abstand von mindestens 10\AA vom Komplex zu den Seiten der Box eingehalten wird.

Das weitere Simulationsprotokoll ist Kapitel 3.3.2 zu entnehmen. Die einzige Ausnahme bildet die Simulationslänge der Modellstrukturen, welche hier 50ns betragen.

Signifikanzanalyseprotokoll Dieses Kapitel lehnt sich an das Protokoll aus Kapitel 3.4.2 an. Es werden jedoch nur noch C_α Distanzen verwendet, alle anderen Distanzen und RMSF Werte werden nicht mehr genutzt. Im letzten Kapitel haben sich die anderen Distanzen auf die PDB Konformation als Referenzstruktur bezogen. In diesem Kapitel liegt eine Modellierung ohne Referenzzustand vor. Damit ist eine Untersuchung solcher Distanzen wenig aussagekräftig. Das sprunghafte Erscheinen und Vergehen des RMSF konnte im vorherigen Kapitel nicht eindeutig geklärt werden. Außerdem wird der RMSF immer zu einer Referenz in dem Fenster berechnet (Kapitel 2.6.3), dass ihn somit zu einer fensterspezifischen Größe und nicht zu einer systemspezifischen Größe, macht.

Clusterprotokoll Zum Clustern wird die *cluster* Funktion von CPPTRAJ, implementiert in AmberTools12 [177] unter Verwendung der paarweisen mittleren

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

quadratischen Abweichung (RMSD) verwendet. Für alle verwendeten Zeitfenster wird eine Gesamtzahl von 2,3,4 und 5 Clustern errechnet. Der *average linkage* Algorithmus wird nur Berücksichtigung der C_α Atome eingesetzt. Andere Clusterparameter werden als Standard verwendet. Zur Validierung einer sinnvollen Clustergröße wird der DBI (Davies-Bouldin Index) als eine Funktion der Anzahl von Clustern mittels des Ellbogen Verfahrens aufgetragen [106]. Eine repräsentative Struktur jedes Clusters wird verwendet, um die mutmaßliche Bindungsarten zu analysieren. Diese Struktur ist als diejenige Struktur definiert, die am nächsten zur Mitte des jeweiligen Clusters liegt. Zur Validierung der Bindungsart werden mehrere Zeitfenster analysiert.

3.4.3 Ergebnisse und Diskussion

Variation der Variablen von SAM in PD1R:PI3K-SH3 Kapitel 3.3 zeigt ein Computerprotokoll, dass statistisch repräsentatives Verhalten in Molekulardynamiksimulationen findet. Ein faszinierendes Merkmal des Ansatzes ist die individuelle Behandlung von Datensätzen, die keine Annahmen über das zu untersuchende System erfordern oder intuitive Konvergenzdefinitionen [113] verwenden. Stattdessen wird jeder Datensatz einzeln für sich betrachtet und behandelt. Im Folgenden wird dieser Ansatz nun verfeinert und auf Anzahl der benötigten Trajektorien, Permutationsgrößen und FDR untersucht. Dabei werden nur noch C_α Distanzen untersucht, da sie die einzige unabhängige Größe in Kapitel 3.4.3 sind (siehe Kapitel 3.4.2). Diese Erkenntnisse sind sehr speziell und damit systemabhängig. Daher empfiehlt es sich, diese Untersuchungen bei jedem neuen System durchzuführen, um einen möglichst großen Nutzen aus der Methode ziehen zu können.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

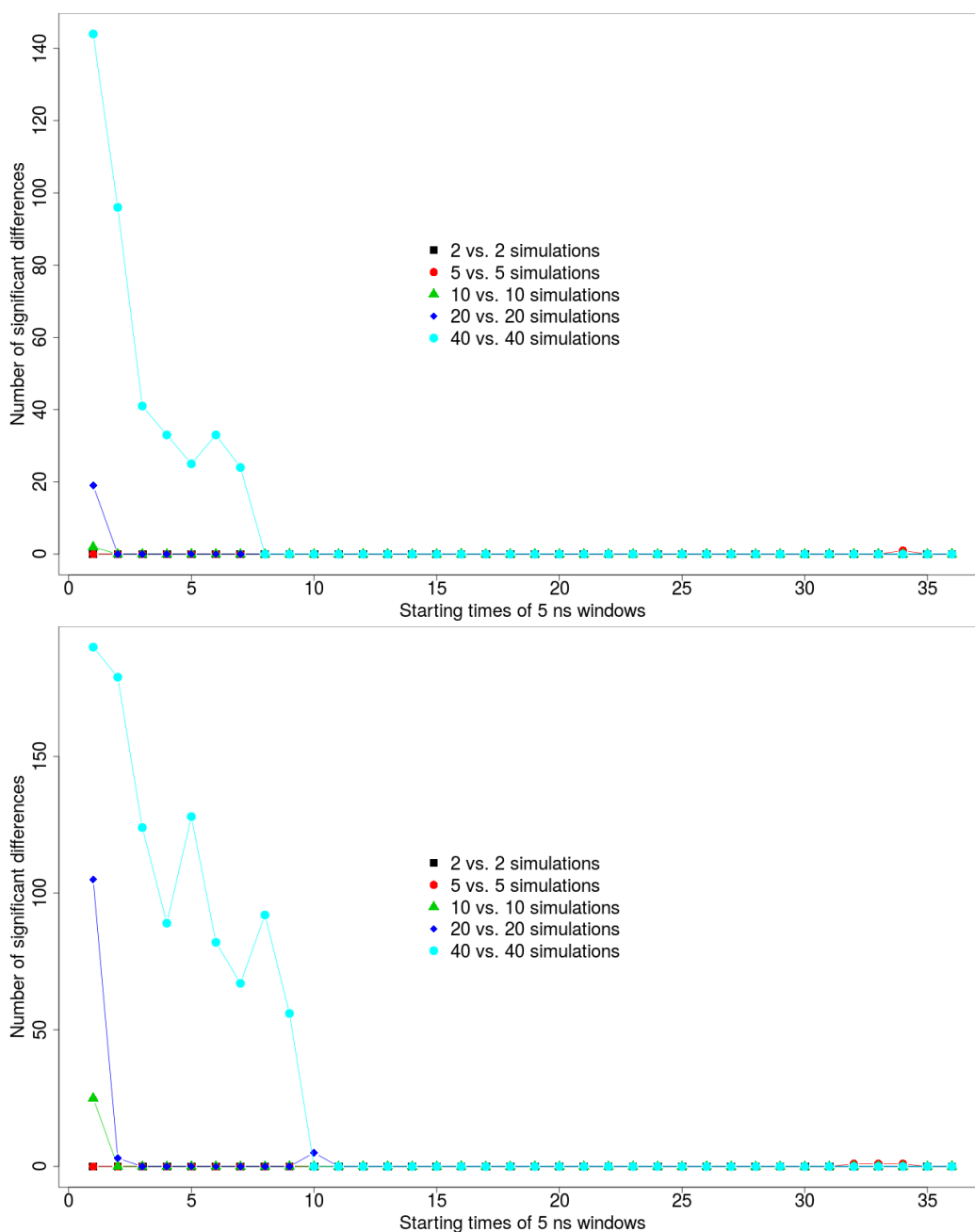


Abbildung 3.26: SAM Analyse von PD1R:PI3K-SH3 mit 100 Permutationen. Die Anzahl der signifikanten Unterschiede (Number of significant differences) wird auf der Y-Achse gezeigt. Die FDR des oberen Graphen beträgt 1 %, die des unteren Graphen 5 %. Die X-Achse beschreibt den Startpunkt eines Zeitfensters in den Trajektorien mit der Länge von 5 ns.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

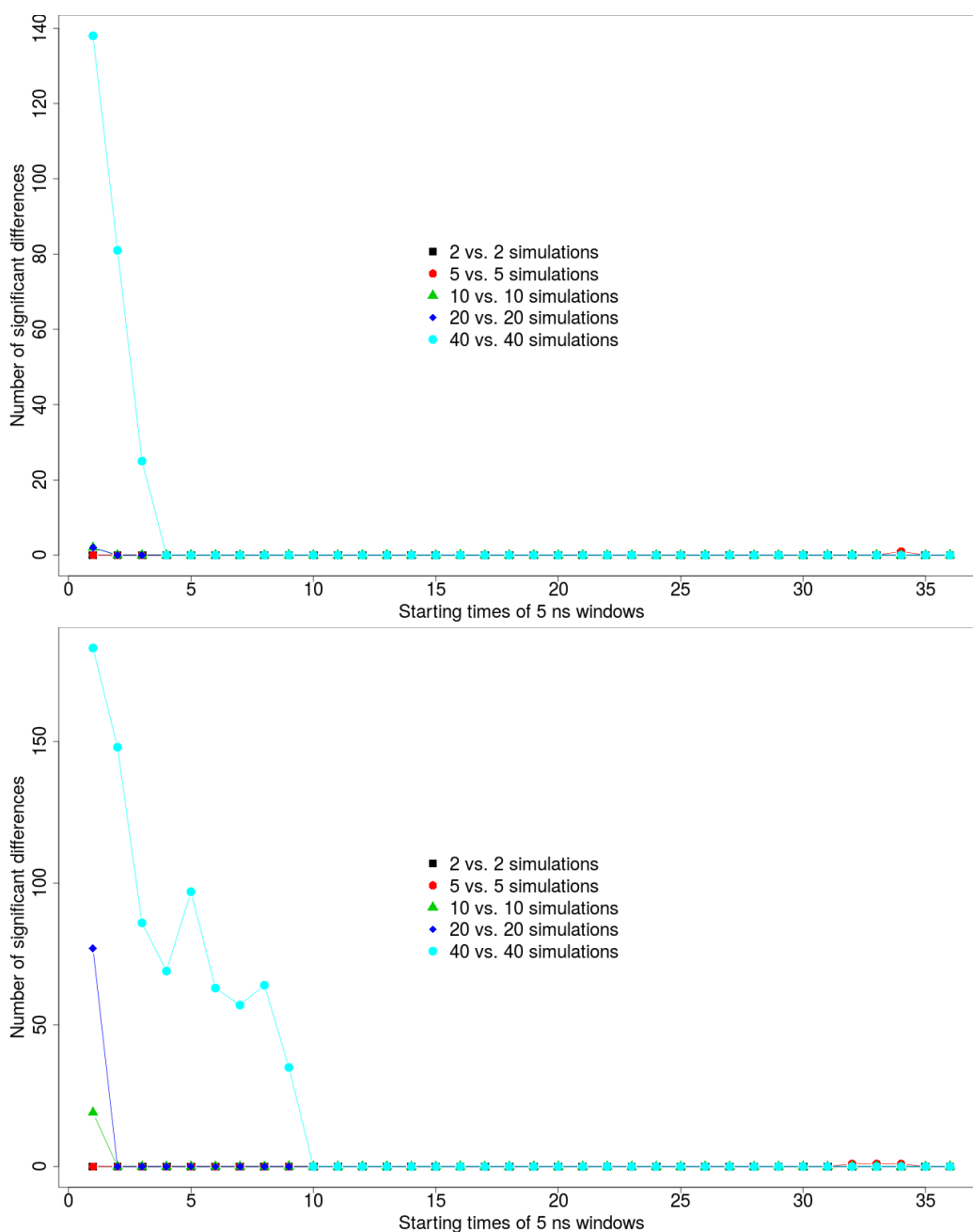


Abbildung 3.27: SAM Analyse von PD1R:PI3K-SH3 mit 1000 Permutationen. Die Anzahl der signifikanten Unterschiede (Number of significant differences) wird auf der Y-Achse gezeigt. Die FDR des oberen Graphen beträgt 1 %, die des unteren Graphen 5 %. Die X-Achse beschreibt den Startpunkt eines Zeitfensters in den Trajektorien mit der Länge von 5 ns.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

Einfluss von Permutationsgröße und FDR In Abbildung 3.26 und 3.27 ist der Einfluss der Permutationsgröße und der FDR für PD1R:PI3K-SH3 gezeigt. Abbildung 3.26 zeigt dabei eine Permutationsgröße von 100, wobei Abbildung 3.27 eine Permutationsgröße von 1000 beschreibt. In beiden Abbildungen ist im oberen Graphen eine FDR von 1 %, und im unteren Graphen eine FDR von 5 % gezeigt. Zusätzlich zeigen alle vier Graphen neben einer Analyse mit allen Trajektorien (40 vs. 40 *simulations*), auch Analysen mit 20, 10, 5 und 2 Trajektorien. Wenn nachfolgend nicht anders erwähnt, beziehen sich die Aussagen auf den Vergleich von 40 gegen 40 Trajektorien. Bei einem Blick auf alle Graphen fällt direkt auf, dass alle Kurven einem Abwärtstrend folgen und früher oder später keine signifikanten Unterschiede mehr zeigen. Des Weiteren folgen drei der vier Graphen einem ähnlichen Kurvenverlauf. Ein interner Vergleich der Abbildungen zeigt in beiden Fällen mehr signifikante Unterschiede in den unteren Graphen, welche eine FDR von 5 % zeigen. Außerdem halten die signifikanten Unterschiede in den 5 % Graphen länger an, und zwar einheitlich bis zum zehnten Zeitfenster. Bei einer strengeren FDR, hier 1 %, ist dies nicht der Fall. Hier scheint die Permutationsgröße einen Einfluss auf die Länge der signifikanten Unterschiede zu haben. Somit kann man methodisch folgern, dass die Permutationsgröße und die FDR ab einem gewissen Grad gegenläufig sind. Eine hohe Permutationsgröße und eine strenge FDR drücken die signifikanten Unterschiede aus, wohingegen eine niedrige Permutationsgröße und eine milde oder moderate FDR die signifikanten Unterschiede ansteigen lassen. Strukturell lässt sich folgern, dass nach einer gewissen Simulationszeit keine signifikanten Unterschiede mehr zu erkennen sind. Damit lässt sich ab diesem Punkt auch keine Startstruktur mehr identifizieren. Simulationen eines Modells sind ab diesem Zeitpunkt genauso aussagekräftig wie Simulationen einer experimentell verifizierten Struktur. Eventuell modellierte Fehler, wie hier einer Aminosäure, die als Startstruktur nicht in die Kompass tasche zeigt, lösen sich damit nach einer gewissen Zeit auf. SAM zeigt den Zeitpunkt der Auflösung an.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

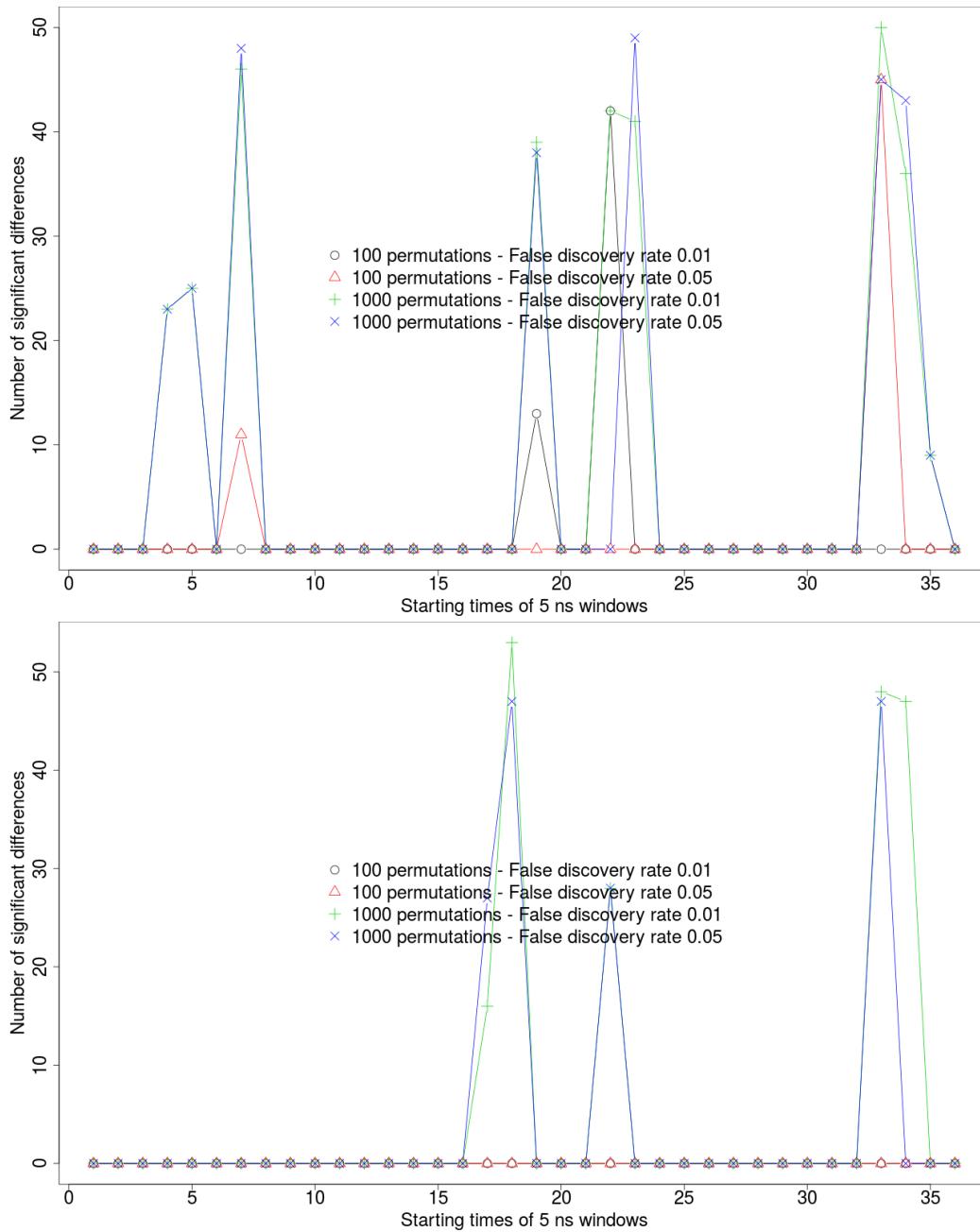


Abbildung 3.28: Analyse des Grundrauschens in PD1R:PI3K-SH3. Untersucht werden 20 gegen 20 Simulationen innerhalb eines Rotamers als Startstruktur. Im oberen Graphen wird die Referenzstruktur untersucht, im unteren die Modellstruktur. Die Anzahl der signifikanten Unterschiede (Number of significant differences) wird auf der Y-Achse gezeigt. Die X-Achse beschreibt den Startpunkt eines Zeitfensters in den Trajektorien mit der Länge von 5 ns.

Grundrauschen in PD1R:PI3K-SH3 Grundsätzlich stellt sich bei diesen Arten von Analysen die Frage eines Grundrauschens. In diesem Zusammenhang kann ein System von sich aus Fluktuation zeigen, die nicht mit wichtigen strukturellen Unterschieden einhergehen. Die Abbildungen 3.26 und 3.27 zeigen vereinzelt minimale Unterschiede, z.B. das zehnte Zeitfenster im unteren Graphen der Abbildung 3.26. Zur Klärung solcher Erscheinungen soll nun ein mögliches Grundrauschen untersucht werden. Dazu werden die 40 Simulationen einer Startstruktur in zwei Gruppen aufgeteilt und mit SAM analysiert. Das Ergebnis wird in Abbildung 3.28 gezeigt, wobei ein Vergleich der Simulationen mit der Referenzstruktur im oberen Graphen dargestellt sind und der Modellstruktur im unteren. Die gefundenen Ergebnisse sind dabei unabhängig von der Startstruktur und zeigen ein Grundrauschen von etwa 50 signifikanten Unterschieden. Bei einer näheren Betrachtung fällt auf, dass die signifikanten Unterschiede bei 1000 Permutationen höher sind als bei 100 Permutation. Das ist interessant, da die signifikanten Unterschiede bei dem Vergleich der Rotamersimulationen (Abbildungen 3.26, 3.27) bei einer größeren Anzahl von Permutation weniger werden. Die FDR scheint hingegen weniger Einfluss auf die signifikanten Unterschiede zu nehmen.

SAM Analyse ohne Referenzstruktur Es konnte in zwei Fällen gezeigt werden, dass die SAM Analyse sehr nützlich ist, um Trajektorien statistisch zu überprüfen, und damit biologisch repräsentatives Verhalten zu identifizieren. Daher kann im Folgenden davon ausgegangen werden, dass dies auch für unbekannte Strukturen, bzw. Strukturen ohne Referenz möglich ist. In einem solchen Fall werden, anstatt einer Referenzstruktur und einer Modellstruktur, zwei unterschiedliche Modellstrukturen verwendet. Bei einem Vergleich der beiden Datensätze mit SAM ist davon auszugehen, dass sich biologisch repräsentatives Verhalten dann zeigt, wenn die Datensätze keine signifikanten Unterschiede zeigen.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

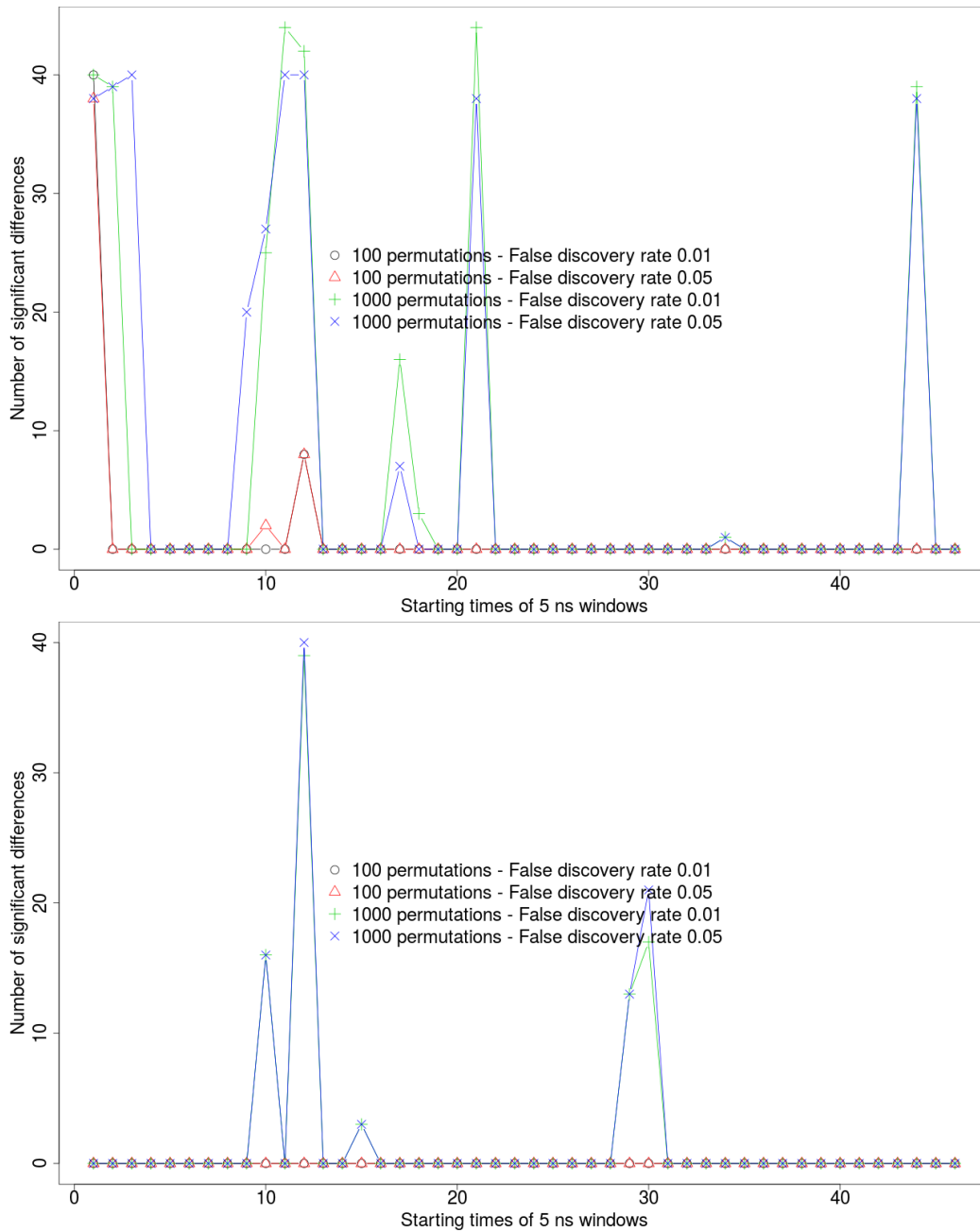


Abbildung 3.29: Analyse des Grundrauschens in PD1R:Hck-SH3. Untersucht werden 20 gegen 20 Simulationen innerhalb eines Rotamers als Startstruktur. Jeder der beiden Graphen untersucht jeweils ein Rotamer der beiden Startstrukturen (linke Abbildung 3.24). Die Anzahl der signifikanten Unterschiede (Number of significant differences) wird auf der Y-Achse gezeigt. Die X-Achse beschreibt den Startpunkt eines Zeitfensters in den Trajektorien mit der Länge von 5 ns.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

SAM Analyse für PD1R:Hck-SH3 Wie im vorherigen Kapitel gezeigt, ist es sinnvoll, zuerst das Grundrauschen der SAM Analyse im Zusammenhang mit einem System zu bestimmen. Sie ist in Abbildung 3.29 gezeigt und folgt dem gleichen Aufbau wie in Abbildung 3.28. Beide Rotamere folgen dabei einem Grundrauschen von etwa 40 signifikanten Unterschieden. Eine Unterscheidung der Permutationsgröße zeigt dabei, ebenso wie im Fall von PD1R:PI3K-SH3, für 1000 Permutationen ein höheres Grundrauschen. Bei einer Permutationsgröße von 100 zeigt sich zwar im ersten Fenster des Rotameres des oberen Graphen ebenfalls ein Grundrauschen von etwa 40 signifikanten Unterschieden, die aber in den nachfolgenden Fenstern nicht wieder erreicht werden. Für das Rotamer im unteren Graphen wird sogar kein Grundrauschen festgestellt. Diese Ergebnisse fließen nun in die nachfolgenden SAM Analysen zwischen den Rotameren mit ein.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

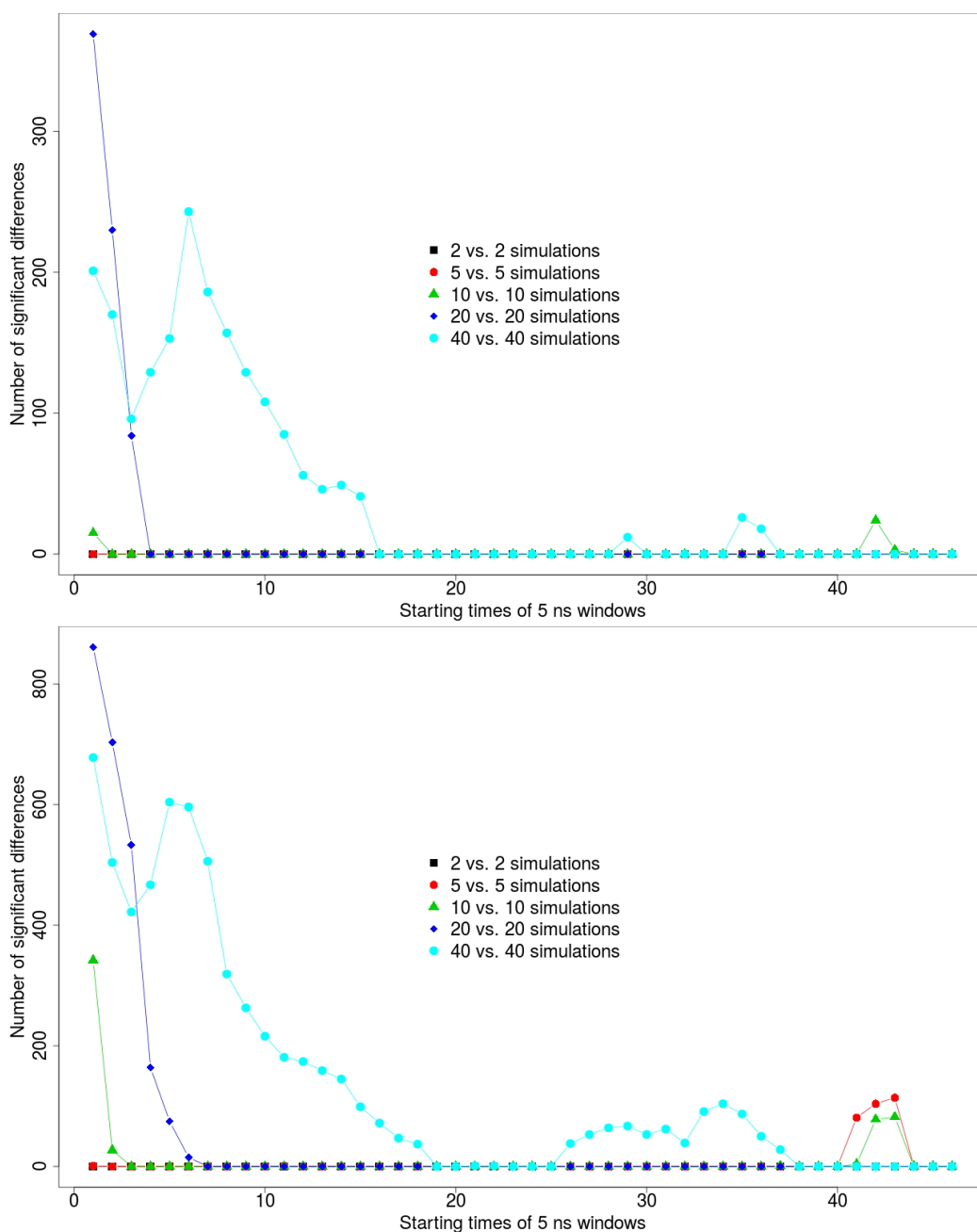


Abbildung 3.30: SAM Analyse von PD1R:Hck-SH3 mit 100 Permutationen. Die Anzahl der signifikanten Unterschiede (Number of significant differences) wird auf der Y-Achse gezeigt. Die FDR des oberen Graphen beträgt 1 %, die des unteren Graphen 5 %. Die X-Achse beschreibt den Startpunkt eines Zeitfensters in den Trajektorien mit der Länge von 5 ns.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

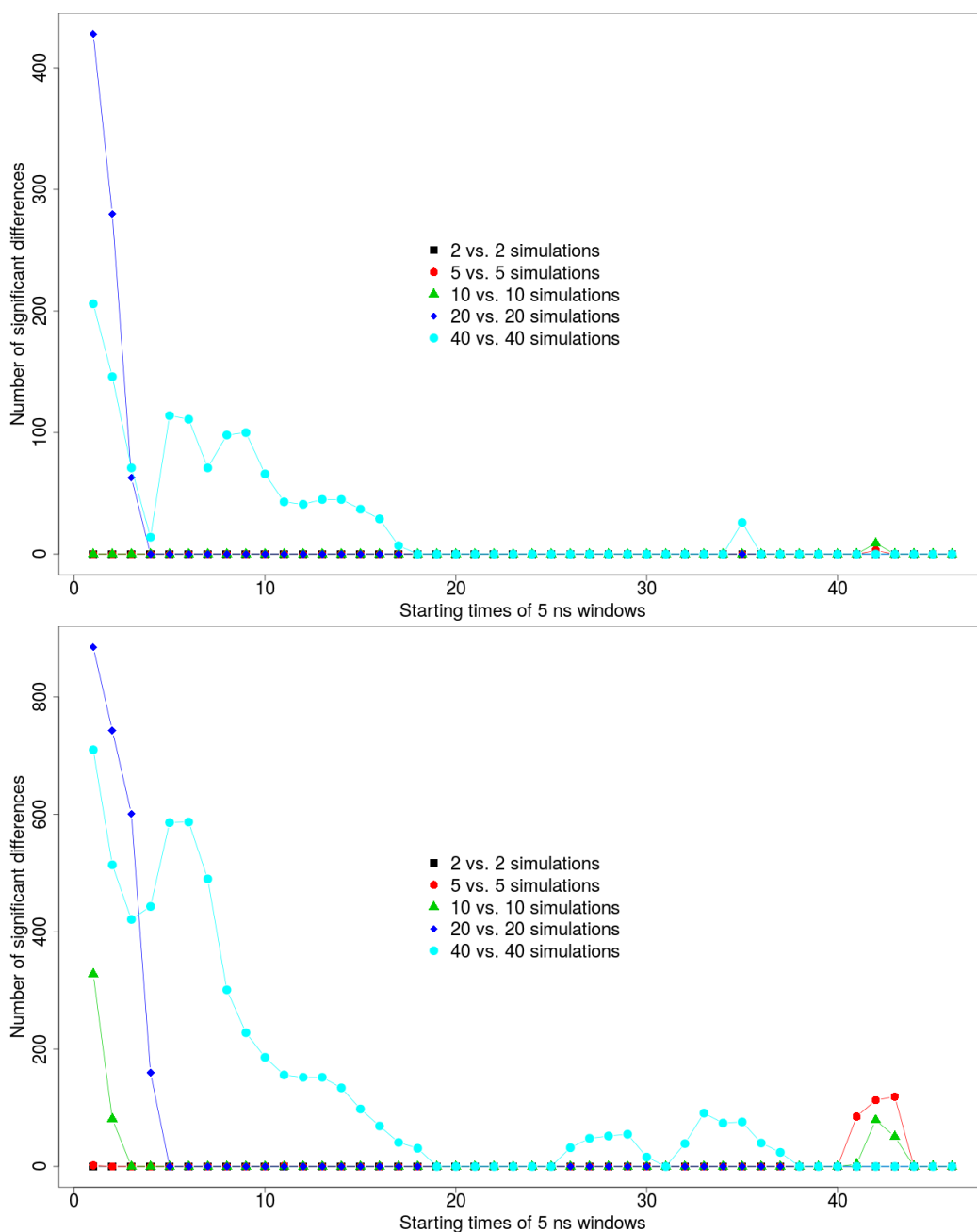


Abbildung 3.31: SAM Analyse von PD1R:Hck-SH3 mit 1000 Permutationen. Die Anzahl der signifikanten Unterschiede (Number of significant differences) wird auf der Y-Achse gezeigt. Die False Discovery Rate des oberen Graphen beträgt 1 %, die des unteren Graphen 5 %. Die X-Achse beschreibt den Startpunkt eines Zeitfensters in den Trajektorien mit der Länge von 5 ns.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

Die Abbildungen 3.30 und 3.31 fassen die Ergebnisse dieser Analysen zusammen. Analog zur Analyse von PD1R:PI3K-SH3 zeigen die jeweils oberen Graphen eine FDR von 1 % und die unteren Graphen eine FDR von 5 %. Die Abbildungen selbst beziehen sich hingegen wieder auf die Permutationsgrößen von 100 für Abbildung 3.30, und 1000 für Abbildung 3.31. Alle vier Graphen zeigen neben einer Analyse mit allen Trajektorien (40 vs. 40 *simulations*), auch Analysen mit 20, 10, 5 und 2 Trajektorien. Wenn nachfolgend nicht gegenteiliges ausgesagt wird, beziehen sich die Aussagen auf den Vergleich von 40 gegen 40 Trajektorien. Grundsätzlich zeigen alle Graphen ein Auslaufen der signifikanten Unterschiede innerhalb der ersten 16 – 19 Zeitfenster. Strukturell kann daraus ein starker Einfluss der initialen Punktmutation (Tyr4Arg) auf die ursprünglichen Konfigurationen des Systems abgeleitet werden. Am Ende treffen sich beide Startstrukturen in dem gleichen Satz von Konfigurationen. Ob diese dann von den ursprünglichen Konfigurationen abweichen, wird weiter unten in diesem Kapitel geklärt. Unter Berücksichtigung des Grundrauschens wird ein Auslaufen dieser Unterschiede um jeweils ein bis zwei Fenster vorgezogen. Allen Graphen zeigen eine wiederkehrende Welle von signifikanten Unterschieden um das 29 und 35 Zeitfenster herum. Bei einer Berücksichtigung des Grundrauschens sind diese Unterschiede aber bei einer FDR von 1 % nicht mehr feststellbar. Leider ist dies bei einer FDR von 5 % nicht der Fall. Somit stellt sich die Frage ob diese Welle in strukturellen Änderungen resultiert. Da im nächsten Arbeitsschritt (Abbildung 3.23) einzelne Fenster ohne signifikante Unterschiede verwendet werden, werden hier Fenster sowohl vor als auch nach der Welle analysiert und miteinander verglichen.

Interessanterweise zeigt ausschließlich der Vergleich von 20 gegen 20 Trajektorien keine späteren Wellen. Der Vergleich von 2 gegen 2 Trajektorien wird dabei vernachlässigt da er niemals Differenzen zeigt, und somit vermutlich ein zu kleiner Datensatz für solche Rechnungen ist.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

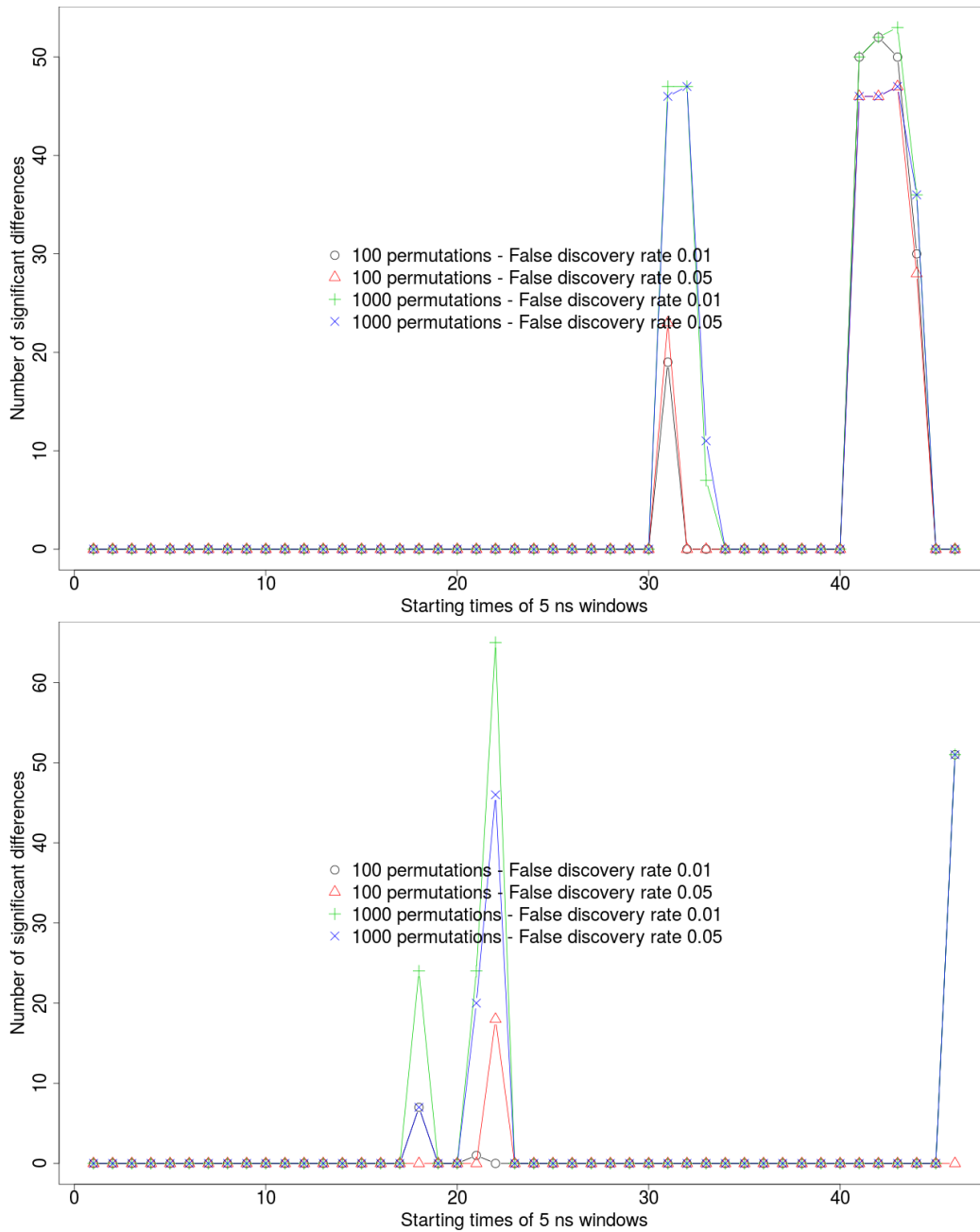


Abbildung 3.32: Analyse des Grundrauschens in PD1:PI3K-SH3. Untersucht werden 20 gegen 20 Simulationen innerhalb eines Rotamers als Startstruktur. Jeder der beiden Graphen untersucht jeweils ein Rotamer der beiden Startstrukturen (rechte Abbildung 3.24). Die Anzahl der signifikanten Unterschiede (Number of significant differences) wird auf der Y-Achse gezeigt. Die X-Achse beschreibt den Startpunkt eines Zeitfensters in den Trajektorien mit der Länge von 5 ns.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

SAM Analyse für PD1:PI3K-SH3 Analog zur statistischen Analyse der Simulationen von PD1R:Hck-SH3 findet hier als erstes eine Analyse des Grundrauschens statt. Die Ergebnisse werden wiederum in zwei verschiedenen Graphen, einer für jede Startstruktur, der Abbildung 3.32 dargestellt. Sie zeigt mit Ausnahme eines Ausreißers bei 60 signifikanten Unterschieden ein Grundrauschen von 50 signifikanten Unterschieden.

In den SAM Analysen der Rotamere lässt sich ein einheitliches Bild darstellen. Die Ergebnisse sind in Abbildung 3.33 und 3.34 zusammengefasst. Auch hier laufen die signifikanten Unterschiede im Laufe der Simulationszeit aus, hier aber im Bereich des 10 – 12 Zeitfensters aus. Unter Berücksichtigung des Grundrauschens verkürzen sich die Zeiten im Mittel auf die Hälfte. Vereinzelte nachfolgende Peaks von signifikanten Unterschieden fallen unter die Schwelle des Grundrauschens.

Interessanterweise zeigen alle Graphen signifikante Differenzen in den Datensätzen von 10 gegen 10 Trajektorien, die aber nicht von den anderen Datensätzen reproduziert werden. Da weitere signifikante Differenzen nur im 40 gegen 40 Trajektorien Datensatz zu finden sind, liegt die Vermutung nahe, dass diese Größe von Datensätzen benötigt wird, um die Entwicklung des Systems nach einer Störung, wie hier durch eine Punktmutation, beobachten zu können.

Allgemein fällt in diesem Zusammenhang auf, dass nach der Hälfte der Simulationszeit keine neuen Erkenntnisse mehr gewonnen werden können. Dieses Ergebnis unterstützt die in der Einführung diskutierten Zusammenhänge zwischen Simulationslänge und mehreren Simulationen mit unterschiedlichen Startbedingungen.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

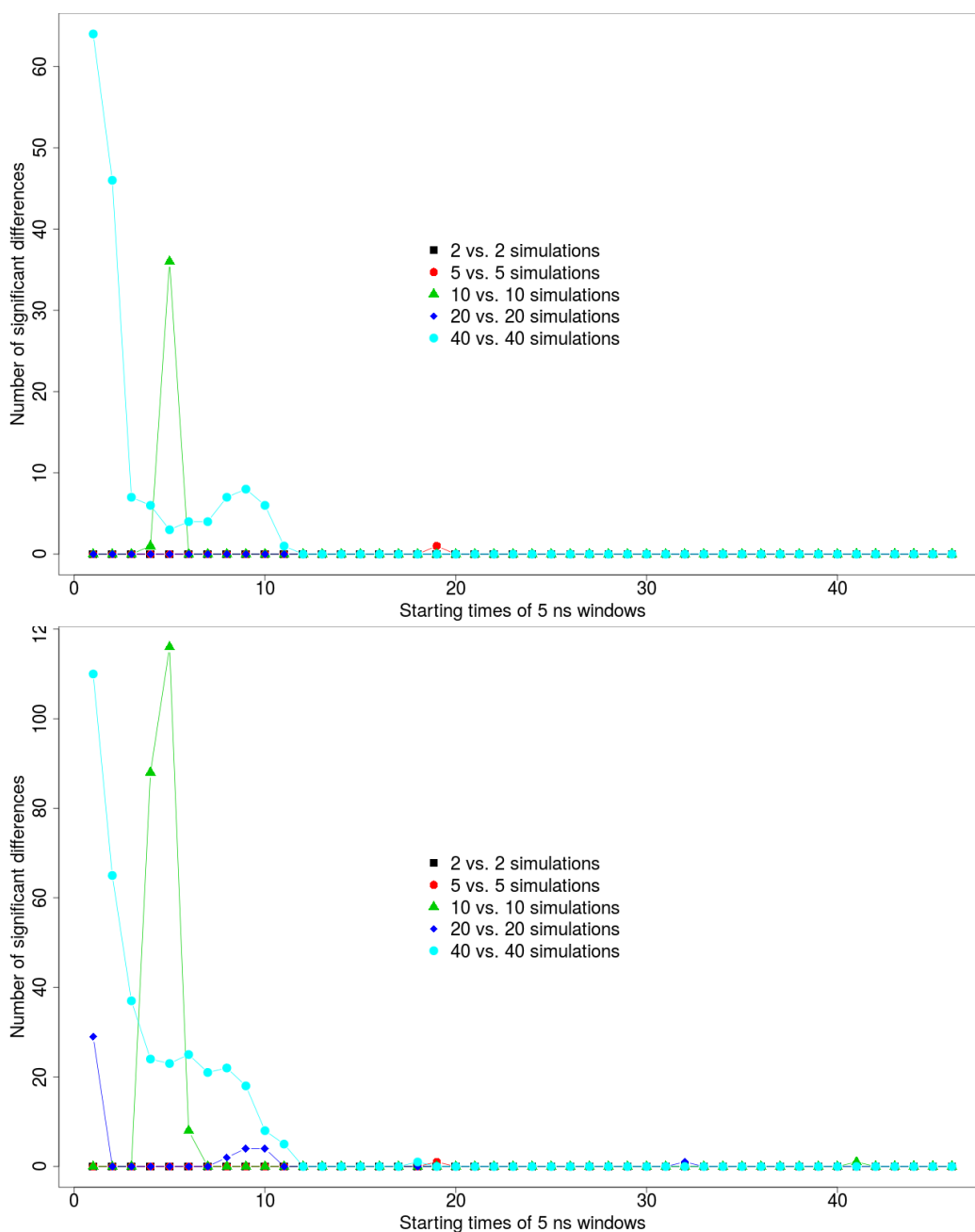


Abbildung 3.33: SAM Analyse von PD1:PI3K-SH3 mit 100 Permutationen. Die Anzahl der signifikanten Unterschiede (Number of significant differences) wird auf der Y-Achse gezeigt. Die FDR des oberen Graphen beträgt 1 %, die des unteren Graphen 5 %. Die X-Achse beschreibt den Startpunkt eines Zeitfensters in den Trajektorien mit der Länge von 5 ns.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

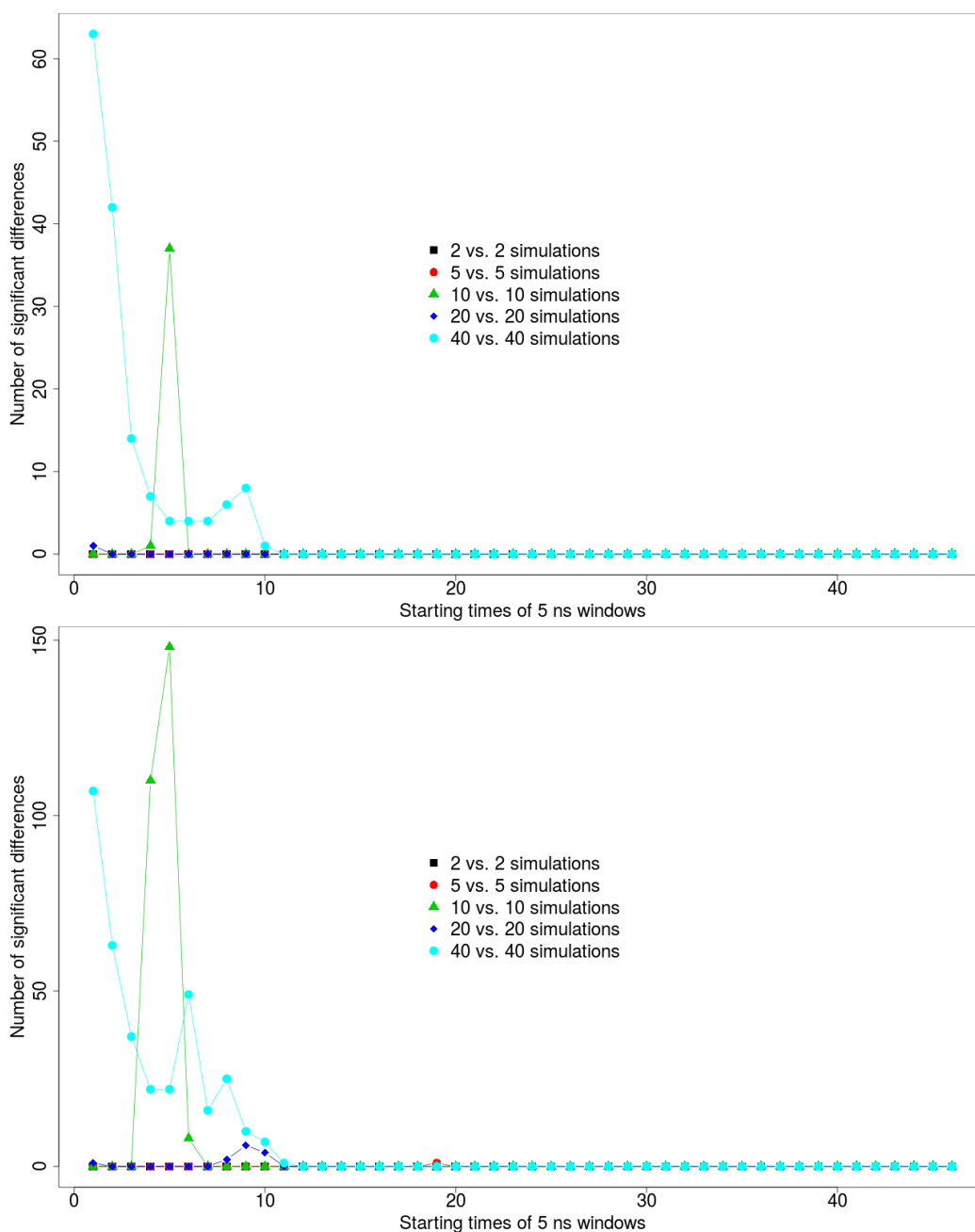


Abbildung 3.34: SAM Analyse von PD1:PI3K-SH3 mit 1000 Permutationen. Die Anzahl der signifikanten Unterschiede (Number of significant differences) wird auf der Y-Achse gezeigt. Die FDR des oberen Graphen beträgt 1 %, die des unteren Graphen 5 %. Die X-Achse beschreibt den Startpunkt eines Zeitfensters in den Trajektorien mit der Länge von 5 ns.

Strukturanalyse der Zeitfenster Basierend auf den vorherigen Ergebnissen werden die Zeitfenster 20, 29, 40 und 45 ausgewählt, um ein Modell für PD1R:Hck-SH3 zu beschreiben. Analog werden für ein Modell von PD1:PI3K-SH3 die beiden Zeitfenster 33 und 41 untersucht. Diese Zeitfenster werden für ein Clustering benutzt, dass repräsentative Strukturen für jeden Cluster ausgibt (Kapitel 2.6.4). Jedes Clustering zeigt dabei einen Hauptcluster aus der mindestens 80 % für PD1R:Hck-SH3, bzw. mindestens 97,5 %, für PD1:Hck-SH3 der berechneten Strukturen aus allen Trajektorien enthält. Die Anzahl der zu berechnenden Cluster wurde dabei für jedes Zeitfenster separat durch die Ellbogen Methode [106] berechnet. Nachfolgend werden die repräsentativen Strukturen mit einer Analyse der Sekundärstruktur untersucht. Hierzu werden die Proteinrückgrat RMSD Werte der Sekundärstruktur (Abbildung 1.4) mit beiden Referenzstrukturen zur Modellierung verglichen. Dabei wird die Einteilung der Sekundärstruktur von Larson und Davidson [38] übernommen. Die Ergebnisse zeigen Tabelle 3.12. Ein gelbes für PD1R:Hck-SH3, oder grünes für PD1:PI3K-SH3 Feld bedeuten, dass das Modell ähnlich seiner Referenzstruktur zur Modellierung ist. Ein rotes Feld markiert hingegen, dass das Modell ähnlich der Referenzstruktur ist, aus der das Modell nicht entwickelt wurde, mit welcher es sich aber die gleiche Sequenz des Peptidliganden teilt. Beispielsweise wird der erste Cluster des 20 Zeitfensters in der Modellstruktur PD1R:Hck-SH3 für die RT-Schleife rot gezeigt. Dies beschreibt, via RMSD, eine höhere strukturelle Ähnlichkeit zu PD1R:PI3K-SH3 als zur entsprechenden Referenzstruktur der Modellierung PD1:Hck-SH3. Diese Ergebnisse werden darüber hinaus mit dem gleichen Farbschema wie in Abbildung 3.35 gezeigt. Die obere Struktur der Abbildung zeigt dabei die Sekundärstruktur von PD1:Hck-SH3. Dies ist die Ausgangsstruktur von PD1R:Hck-SH3 ist. Die untere Struktur zeigt PD1R:PI3K-SH3. Dies ist die Ausgangsstruktur von PD1:PI3K-SH3 ist. Für die Abbildung 3.35 muss angemerkt werden, dass die Rotfärbung der Segmente eine Zusammenfassung für alle roten Segmente in Tabelle 3.12 ist.

PD1R:Hck-SH3 Die Ergebnisse für PD1R:Hck-SH3, aus Tabelle 3.12 und der oberen Abbildung 3.35 sind unerwartet, da sie zeigen, dass acht der elf Sekundärstrukturelemente ihre Konformation verändert haben. Dies bedeutet, dass das Modell für PD1R:Hck-SH3 einer Struktur mit gleichem Ligand ähnlicher ist wie

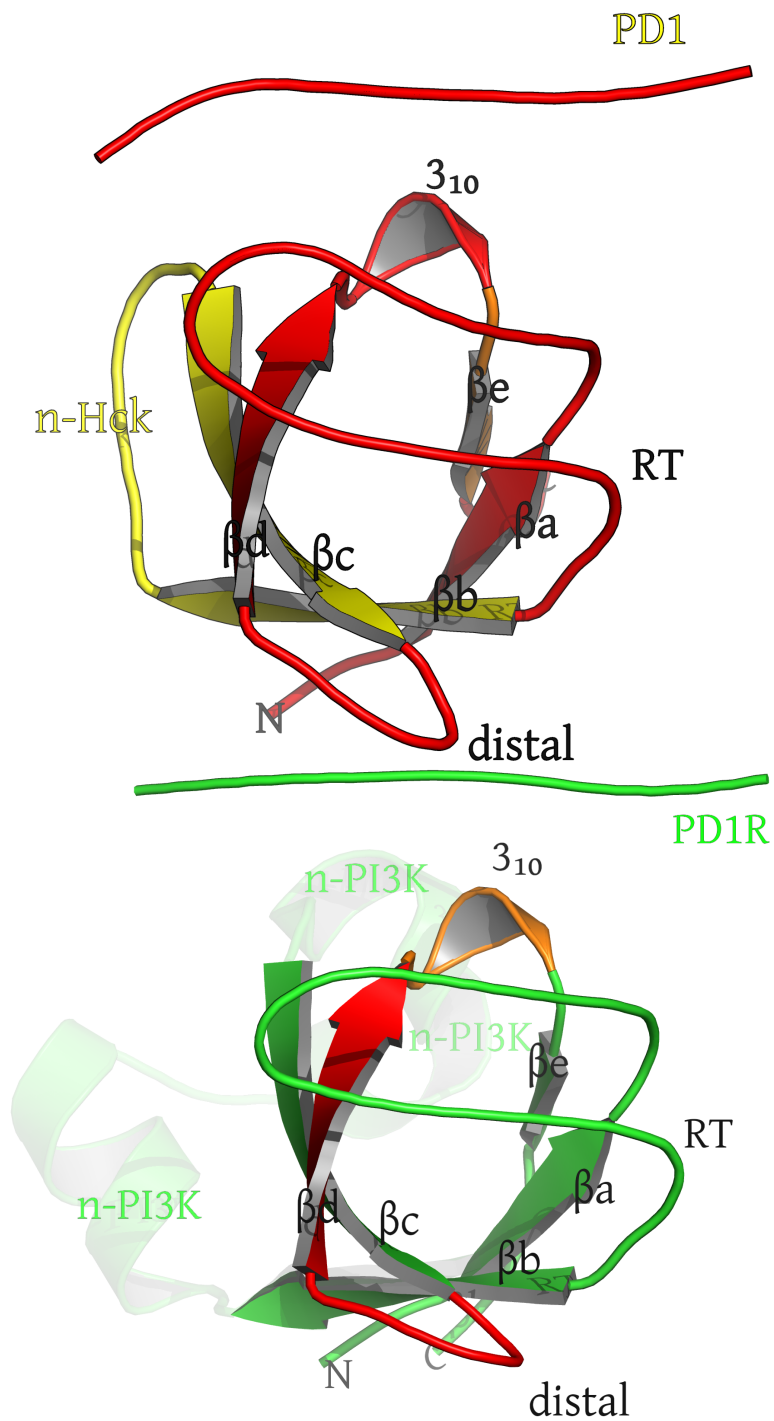


Abbildung 3.35: Sekundärstruktur von PD1:Hck-SH3 (gelb (OBEN)) und PD1R:PI3K-SH3 (grün (UNTEN)). Segmente, die nach der Simulation eine andere Konformationen eingenommen haben, werden dabei rot angezeigt, analog zur Tabelle 3.12.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

seine Referenzstruktur aus der Modellierung.

Wie bereits erwähnt, wird auch das Zeitfenster 29, das in drei von vier Fällen (Abbildungen 3.30, 3.31) signifikante Unterschiede zeigt, als Vergleich ausgewertet. Der wichtigste Unterschied zu den anderen Zeitfenstern zeigt sich hier in der fehlenden Konformationsänderung des β d-Faltblattes (Tabelle 3.12). Wahrscheinlich ist somit, dass die zweite Welle in den Abbildungen 3.30 und 3.31 verschiedene Konformationsänderungen in der Mitte der SH3 Domäne anzeigt, die dann in einer Konformationsänderung der distalen-Schleife resultieren kann. Die Konformation des Liganden ist hingegen in allen Clustern gleich. Ein genauerer Blick auf die atomare Ebene von PD1R:Hck-SH3 (Abbildung 3.21 unten) zeigt, dass der beim Starten der Simulation vorhandene Knick der Referenzstruktur in dem Modell verschwunden ist. Solche Effekte können somit als Modellierungsfehler betrachtet werden. Eine Auswertung der Trajektorien an der falschen Stelle könnten dadurch zu falschen Schlussfolgerungen führen. In diesen Strukturen ohne Knick wird nun das modellierte Arg4 in der Kompass tasche gefunden. Hierbei ist die Guanidiniumgruppe des Arg4 mit der Carboxylatgruppe von D17

Tabelle 3.12: Vergleich von Sekundärstrukturelementen. Jedes grüne oder gelbe Feld bezieht sich auf ein Segment, das einen RMSD Wert analog zur Referenzstruktur, aus der das Modell generiert wurde, besitzt. Jedes rote Feld bezieht sich auf ein Segment, das einen RMSD Wert ähnlich der Referenzstruktur besitzt, die die gleiche Sequenz wie der Ligand hat.

	PD1R:Hck-SH3								PD1:PI3K-SH3				
Zeitfenster	20			29		40		45	33		41		
Cluster	1	2	3	1	2	1	2	1	2	1	2	1	2
X													
β a-Faltblatt													
RT-Schleife													
β b-Faltblatt													
n-Scr-Schleife													
β c-Faltblatt													
Distal-Schleife													
β d-Faltblatt													
3_{10} -Helix													
β e-Faltblatt													
Peptid													

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

eine Wasserstoffbrückenbindungen eingegangen. Zur Erinnerung, die Kompassstasche wird als Hauptinteraktionspartner für basische Aminosäuren in Peptiden beschrieben [37, 185, 1.2.3]. Dieses Ergebnis ist nur möglich, wenn Lys3 aus der Kompassstasche (Abbildung 3.36 oben) vertrieben wird. In diesem Zusammenhang sind Interaktionspartner für Lys3 nicht mehr vorhanden und die Seitenkette ist zum Lösemittel hin ausgerichtet. Diese Konformation ist schon für Lys3 in PD1R:PI3K-SH3 gefunden worden. Der Knick wird dadurch überflüssig und Wasserstoffbrückenbindungen, die diesen Knick stabilisiert haben [37], fehlen somit nachweislich auch. Seitenketten der n-Scr-Schleife zeigen in diesem Modell nur sporadische Wechselwirkungen mit den ersten vier Aminosäuren des Liganden. Wechselwirkungen zwischen dem hydrophoben Teil des Liganden PD1R (5-12) und Hck-SH3 sind konserviert in Bezug auf die Referenzstruktur [37]. Dies unterstreicht die früheren Untersuchungen, die zwei Arten von Bindungen beschrieben haben und die im Bereich der W36 ineinander übergehen. Einige der kleineren Cluster bieten interessante Wechselwirkungen. Sie zeigen, dass im Falle, in dem Arg4 außerhalb der Kompassstasche liegt, diese Tasche leer bleibt. Dies führt zu instabilen Konformationen im Bereich der ersten vier Aminosäuren in PD1R. Nur der hydrophobe Teil (Aminosäuren 5-12) ist noch gebunden. In einem anderen Fall wird Arg4 über N_δ und N_ω an das D17-carboxylat (Abbildung 3.36 unten) gebunden, wohingegen die Hauptcluster Interaktionen über N_ω' und N_ω (Abbildung 3.36 oben) zeigen. Im Falle der N_δ , N_ω Bindung zeigt Lys3 Wechselwirkung mit E35 in der n-Scr-Schleife, auf der gegenüberliegenden Seite der Bindungstasche. Dadurch wird das Peptidrückgrat im Bereich von Lys3 gestreckt und die Seitenkette von Arg4 kann gebogen werden, so dass die Guanidiniumgruppe auf diese Weise wechselwirken kann. Aus kompensatorischen Gründen ist das Peptidrückgrat im Bereich der Aminosäuren 5-8 in einer Konfiguration außerhalb des klassischen PPII Motives (untere Struktur der Abbildung 3.36).

PD1:PI3K-SH3 Für ein Modell von PD1:PI3K-SH3, werden die Zeitfenster 33 und 41 (Tabelle 3.12), die ohne signifikante Unterschiede (Abbildungen 3.33, 3.34) sind, untersucht. Die Strukturen, die die Hauptcluster aus dieser Untersuchung repräsentieren, haben dabei hohe Ähnlichkeit mit der Referenzstruktur, aus der sie generiert worden sind und weisen deshalb nur wenige rote Segmente in der

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

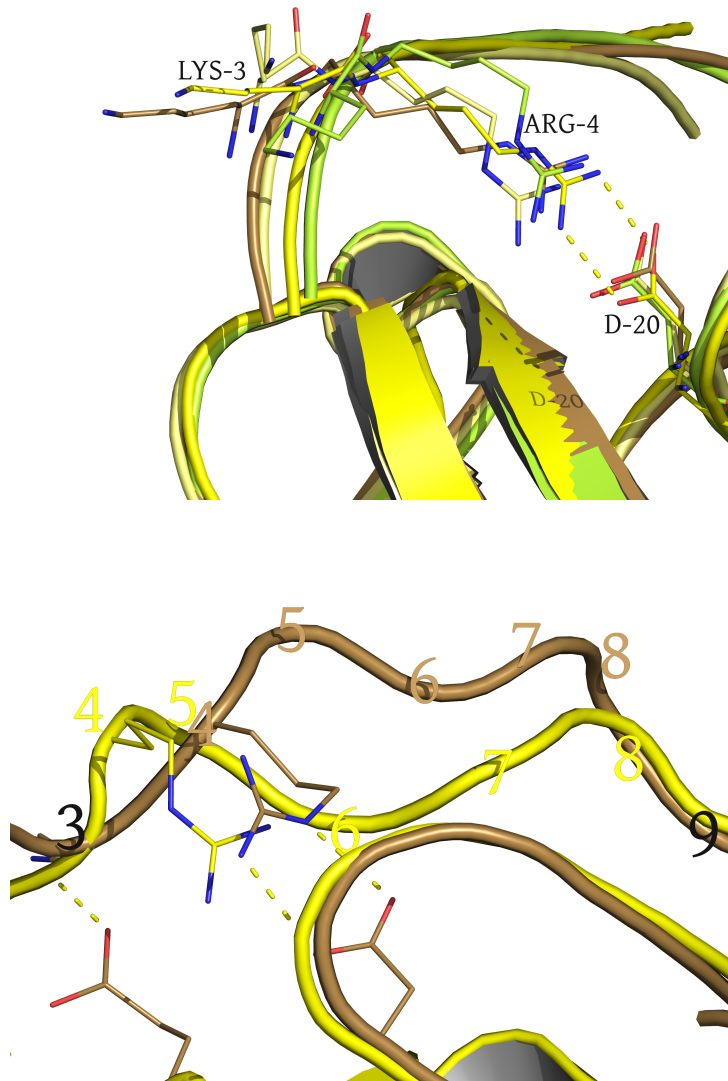


Abbildung 3.36: Nahaufnahme von verschiedenen Bindungsmodi in PD1R:Hck-SH3. OBEN: Vergleich der Hauptcluster aus Tabelle 3.12. Alle Strukturen zeigen die Verdrängung von Lys3 durch Arg4 aus der Bindungstasche. Die am dunkelsten gefärbte Struktur zeigt die Konfiguration im Zeitfenster 29. Das Peptid bindet hier ähnlich wie die Strukturen der anderen Zeitfenster. UNTEN: Struktur aus einem Nebenclustern der Zeitfenster. Sie zeigt einen von den Hauptclustern verschiedenen Bindungsmodus für Arg4 und die Kompensation des Peptidrückgrats.

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

unteren Abbildung 3.35 oder Tabelle 3.12 auf. Diese betreffen ausschließlich das β d-Faltblatt und die vor und nachgeschalteten Schleifen. Abbildung 3.37 zeigt in diesem Zusammenhang detaillierter die Interaktionsfläche zwischen Peptidligand und SH3 Domäne, einmal für die Hauptcluster (OBEN) und die Nebencluster (UNTEN). Die Referenzstruktur zur Modellierung ist dabei gelb dargestellt und die Strukturen aus den Clustern hellbraun und dunkelbraun. Strukturell zeigt die Abbildung in beiden Fällen einen konservierten Zustand der PPII Peptidkonformation. Außerdem ist Lys3 ungebunden ähnlich seiner Referenzstruktur zur Modellierung. Damit nimmt Lys3 nicht die Bindungsstelle von Arg4 in der Kompass tasche ein. Die für beide Zeitfenster gefundenen Nebencluster sind qualitativ ähnlich (untere Abbildung 3.37). Die Peptidliganden haben hier nur durch ihren hydrophoben Teil (Aminosäuren 5-12) Kontakt mit der SH3 Domäne. Außerdem zeigt der klassische PXXP Teil (P_0 , P_3 in Klasse I Liganden) der Peptidliganden nahezu identische Kontakte wie die Referenzstruktur zur Modellierung. Alle Aminosäuren N-terminal hierzu sind in Richtung des Lösemittels orientiert. In diesem Fall können wichtige Wasserstoffbrückenbindungen zwischen dem Sauerstoffatom von Pro8 und Pro5 mit der OH-Gruppe von Y54 und der NH-Gruppe der W36 Indol-Gruppe gefunden werden. Des weiteren können unterstützende Wechselwirkungen zwischen Ser11 und E33, sowie Pro7 und T53 dargestellt werden. In dieser U-förmigen Konformation ist die RT-Schleife weiter von den Faltblättern, die den Kern der SH3 Domäne ausmachen, entfernt (untere Abbildung 3.37). Die Literatur [37, 197] beschreibt W36 als Position, an welcher der hydrophobe Teil eines Peptidliganden in den Kompass teil in Form einer Konformationsänderung übergeht. In den Nebenclustern ist dies die Position, in der der Ligand seine Wechselwirkungen mit der SH3 Domäne verliert und ausschließlich mit seinem hydrophoben Teil (Aminosäuren 5-12) bindet. Diese hydrophoben Wechselwirkungen sind sowohl in den Haupt- als auch in den Nebenclustern gleich den Wechselwirkungen der Referenzstruktur, aus der sie modelliert worden sind. Durch die oben erwähnte RT-Schleife, die vom Rest der SH3 Domäne weg geneigt ist, bekommt Tyr4 mehr Bewegungsspielraum. Eine Analyse der Wasserstoffbrückenbindungen in dieser Region zeigt in 15 % aller Zeitschritte Interaktionen mit den Carboxylatgruppen der Aminosäuren E16 und D17 an der Spitze der RT-Schleife. Bei einem Vergleich der RT-Schleife zwischen Hck- und PI3K-SH3 weisen fünf Aminosäu-

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

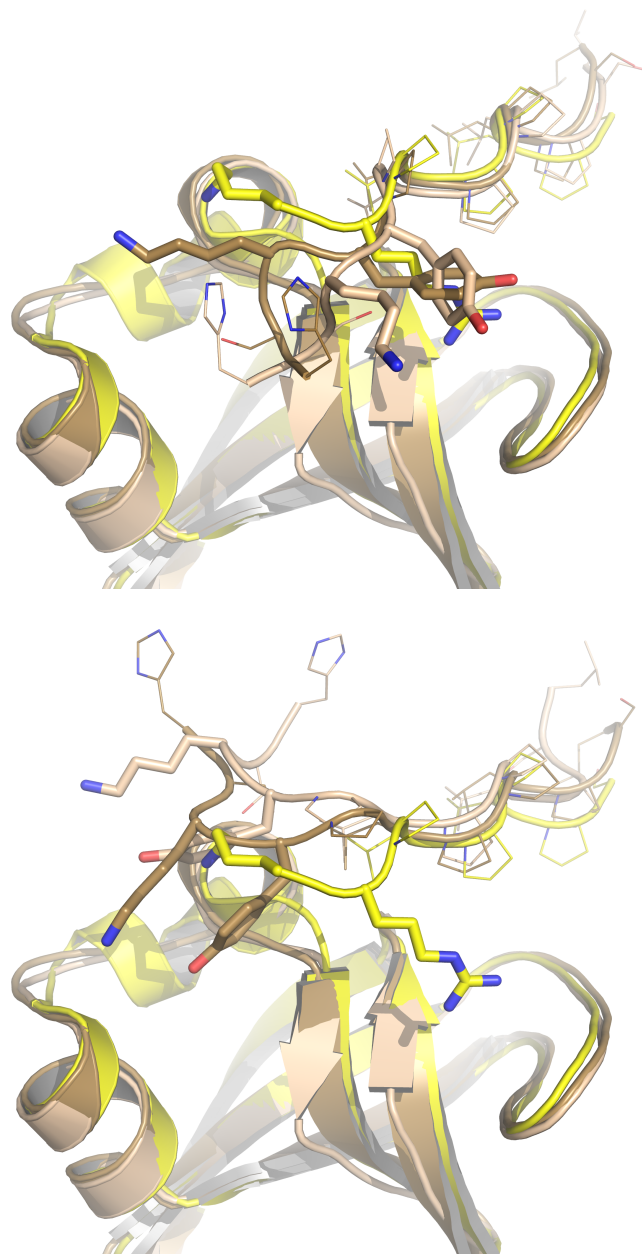


Abbildung 3.37: Nahaufnahme von verschiedenen Bindungsmodi in PD1:PI3K-SH3. In beiden Strukturen ist die Referenzstruktur PD1R:PI3K-SH3, aus der die Modellstrukturen generiert wurden gelb eingezeichnet. Die beiden anderen Strukturen zeigen Strukturen aus den Hauptclustern (OBEN) oder Nebenclustern (UNTEN).

3.4 Modellierung von Bindungsmodi spezifischer Peptid-SH3 Domänen

ren N-terminal zu E16 Unterschiede auf (Tabelle 3.13). Hck-SH3 zeigt vor allem Aminosäuren, die für starke van-der-Waals Wechselwirkungen (H15, H14, I13, A12, E11) bekannt sind. PI3K-SH3 zeigt hingegen geladene Aminosäuren (E15, R14, E13, K12, K11). Tyr4, das einen starken Beitrag zu der hohen Bindungsaffinität in PD1:Hck-SH3 (Tabelle 3.11) über van-der-Waals Wechselwirkungen zeigt, ist daher in diesem Komplex (PD1:PI3K-SH3) nicht möglich. Ein Knick in PD1, welcher von Schmidt et al. [37] für PD1:Hck-SH3 beschrieben wurde, bildet sich hier nicht aus. Dies ist vermutlich auf die leere Kompassstasche zurückzuführen, die hier zu finden ist und bei Schmidt et al. [37] von Lys3 ausgefüllt war. Die Analyse der Wasserstoffbrückenbindungen in dieser Region zeigt Lys3 nur in etwa 6 % aller Zeitschritte in der Kompassstasche koordiniert. Interaktionspartner sind dann E16 und D17, welche auch schon oben bei Interaktionen mit Tyr4 beschrieben worden sind. Da sich der Knick in PD1 nicht ausbildet, sind die dafür nötigen Wasserstoffbrückenbindungen auch nicht zu finden. Sie waren in der Arbeit von Schmidt et al. zwischen Tyr4 und Leu6, sowie zwischen dem Carbonylsauerstoff von Tyr4 und Indolstickstoff von W36 zu finden. In der hier durchgeführten Analyse der Wasserstoffbrückenbindungen zeigen sich gelegentlich Interaktionen von Lys3 mit D49. In Hck-SH3 von Schmidt et al. [37] gab es kein D49 oder eine Alternative. Somit stellt diese Interaktion in PD1:PI3K-SH3 eine systemspezifische Interaktion dar, die sich mit His1 und Ser2 abwechselt.

Tabelle 3.13: Sequenzen der RT-Schleifen von Hck- und PI3K-SH3. Unterschiede zwischen den Sequenzen werden durch Fettdruck gekennzeichnet. Aromatische Aminosäuren werden durch gelben Feldern angezeigt. Die Zählung ist identisch mit der Arbeit von Larson et al. [38].

	RT-Schleife								
	7	8	9	10	11	12	13	14	15
Hck	L	Y	D	Y	E	A	I	H	H
PI3K	L	Y	D	Y	K	K	E	R	E
	16	17	18	19	20	21	22	23	24
Hck	E	D	L	S	F	Q	K	G	D
PI3K	E	D	I	D	L	H	L	G	D

Faltblatt in RT-Schleife Schmidt et al. [37] fand ein kurzes Faltblatt, bestehend aus Y10 und F20 in der RT-Schleife von PD1:Hck-SH3. Das Vorhandensein dieses Faltblattes wird in allen der untersuchten Zeitfenster beider Modellstrukturen bestätigt. In PD1R:Hck-SH3 zeigt sich die Wasserstoffbrückenbindung zwischen der Carbonylgruppe von F20 und der Amidgruppe von Y10 in 70 % aller Zeitschritte. Die entgegengesetzte Wasserstoffbrückenbindung zwischen der Carbonylgruppe von Y10 und der Amidgruppe von F20, ist hingegen nur in 45 % aller Zeitschritte zu finden. In dem Modell PD1:PI3K-SH3, sind beide Wasserstoffbrückenbindungen jeweils in 50 % aller Zeitschritte zu finden. Diese Ergebnisse legen nahe, dass dieses Faltblatt in wahrscheinlich fast allen SH3 Domänen gefunden werden kann. Ein weiterer Grund dafür ist, dass Larson und Davidson diese hydrophoben Positionen konserviert fanden. Dass dieses Faltblatt noch nicht so oft beschrieben wurde, kann daran liegen, dass es in anderen Strukturen vermutlich auch in einer Intensität von nur 50 % vorliegt und somit unbemerkt blieb.

Klasse I' Konfiguration In Kapitel 1.2.3 wurde schon darauf hingewiesen, dass P51 dazu genutzt werden kann, um Liganden zu klassifizieren. Hier soll dieser Zusammenhang näher beschrieben und für die beiden Modellstrukturen untersucht werden. Ein typischer Diederwinkel von Prolin kann in drei Rotamere eingeteilt werden. Er wird meist als χ Winkel bezeichnet. Für das Rotamer 1 ergibt sich somit ein Winkel 1 von <0 und ein Winkel 2 von >0 . Für das Rotamer 2 sind die Verhältnisse umgekehrt und der Winkel 1 ist >0 und der Winkel 2 ist <0 . Für Rotamer 3 liegen beide Winkel bei 0. Wenn eine SH3 Domäne an einen Klasse I Liganden gebunden ist, zeigt sich für P51 eine Rotamer 1 Konformation. Die Gegenwart eines Klasse II Liganden führt zu Diederwinkeln an P51, die für ein Rotamer 2 gelten. Wenn hingegen die Gegenwart eines Liganden mit einem Klasse I Motiv vorliegt und die P51 Konformation in Rotamer 2 Stellung vorliegt, spricht man von einem Klasse I' Liganden [185]. In den Untersuchungen der beiden Modellstrukturen PD1R:Hck-SH3 und PD1:PI3K-SH3 zeigen sich χ Winkel, die einen Klasse I' Liganden beschreiben. Dies bedeutet, dass sie nicht durch die Simulation verschoben werden. Sie zeigen in den Clustern immer noch die gleichen Konformation wie in der Referenzstruktur zur Modellierung.

Ausblick

Molekulardynamiksimulationen werden meistens nach einer sogenannten *equilibration time* analysiert. Nach dieser Zeit wird angenommen, dass ein System nur noch repräsentative Zustände einnimmt und sich im Gleichgewicht befindet. Alle Störungen, die eingeführt werden, sei es durch die Initialisierung (Waterbox, etc.) und / oder die Modellierung der Startstruktur, sind danach nicht mehr wirksam. Dieser Zeitpunkt wird oft willkürlich gesetzt oder von anderen Systemen übernommen. Bei einem solchen Ansatz kann das individuelle Verhalten des Systems jedoch nicht erkannt werden. Die Wahrscheinlichkeit ist groß, dass nicht repräsentative Daten analysiert werden.

Intuitive Konvergenzdefinitionen werden ebenso oft als *equilibration time* verwendet. Streng genommen ist dies unwissenschaftlich [113]. In einem mehr-trajektorien Ansatz von Molekulardynamiksimulationen ist die Analyse aufgrund der Datenmenge auch unpraktisch. Aus diesem Grunde ist ein neuer Ansatz für die Bewertung von Molekulardynamiksimulationen erforderlich ist. Ich habe in dieser Arbeit die SAM Methode auf ihre Brauchbarkeit in diesem Zusammenhang überprüft. Diese und eine andere Arbeit [198] haben gezeigt, dass SAM eine sehr nützliche Methode für die Analyse von Molekulardynamiksimulationen in einem mehr-trajektorien Ansatz ist.

Ähnlich zur SH3 Domäne gibt es weitere Proteine, die als Interaktionsdomänen fungieren, wie zum Beispiel 14-3-3 Proteine [199]. 14-3-3 Proteine sind die ersten Moleküle, von denen berichtet wurde, dass sie in der Lage sind, Phosphoserin- und Phosphothreonin-Motive zu erkennen [200]. Derzeit ist noch nicht vollständig geklärt, warum die unterschiedlichen Isoformen der 14-3-3 Proteine verschiedene Bindungsspezifitäten zeigen. Außerdem existieren noch offene Fragen über die Details der strukturellen Modulation bei der Bindung von Liganden. Beides sind ideale Fragestellungen für das in dieser Arbeit entwickelte Verfahren. Vor allem da hoch aufgelöste Strukturen vorhanden sind, welche eine Auflösung von unter 2 Å besitzen und somit für Molekulardynamiksimulationen geradezu prädestiniert sind [201, 202, 203].

Obwohl zu diesem Zeitpunkt noch nicht alle methodischen Fragen geklärt sind, kann angenommen werden, dass SAM in den nächsten Jahren noch viele Mole-

kulardynamiksimulationen erfolgreich analysieren wird. Beispielsweise könnten drei oder mehr Ausgangsstrukturen getestet werden. Auch könnten Parameter der Methode selbst variiert werden. Mit SAM könnten des Weiteren auch WT gegen MT Simulationen verglichen werden. Hierbei sind die Unterschiede, die durch die Mutation induziert werden, direkt sichtbar, denn jeder signifikante Unterschied ist in der Struktur direkt lokalisierbar. Im Zusammenhang mit den untersuchten SH3 Domänen dieser Arbeit wäre auch interessant, andere Systeme von Tran et al. [184] oder Mutationen in der RT- und n-Src-Schleife zu analysieren. Mutationen in der n-Src-Schleife sind dafür bekannt, dass sie einen großen Einfluss auf die Peptidbindung [185] haben. SAM hat das Potenzial solche Fragen zu beantworten. Um Rechenzeit zu sparen wäre es möglich, Molekulardynamiksimulationen unterbrochen von SAM Analysen zu verwenden. Grundsätzlich kann jeder Parameter bei diesem Verfahren verwendet werden. Einzige Voraussetzung ist die Empfindlichkeit der Parameter und die Menge der Messdaten. Beispielsweise ist es sinnvoll eine erhöhte Anzahl der C α Distanzen zu untersuchen. Auf der anderen Seite ist es nicht sinnvoll, gleiche Distanzen zu untersuchen, die sich gegenseitig implizieren. Als weiterer Parameter könnten verschiedene Arten von Energie verwendet werden. Im besten Fall sind die Parameter nicht starr oder vordefiniert, wie im Fall der Wasserstoffbrückenbindungen oder der van-der-Waals Distanzen (Kapitel 3.4.3). Falls man beabsichtigt, zu einer bestimmten Referenzstruktur zu analysieren, stellt dies kein Problem dar. Dies ist jedoch häufig nicht der Fall. Eine neue Entwicklung des Systems bleibt somit ungesehen. Darüber hinaus zeigen meine Ergebnisse, dass es nicht sinnvoll ist, eine *equilibration time* im Vorhinein zu definieren, da die Zeit, bis sich ein System im Gleichgewicht befindet, individuell ist. Darüber hinaus sind die Systeme der GDH oder Chitinase, die im Laufe dieser Arbeit zusätzlich modelliert wurden, leicht mithilfe von Molekulardynamiksimulationen und SAM erweiterbar. Molekulardynamiksimulationen würden die Modelle vertieft auf ihre Gültigkeit überprüfen. Neue Entwicklungen und wichtige Konformationen würden dann, wie im Fall der SH3 Domäne, mittels Signifikanz- und Clusteranalyse extrahiert werden.

Literatur

- [1] Friedman, R., Boye, K., Flatmark, K., 2013, Molecular modelling and simulations in cancer research., *Biochimica et biophysica acta*, 1836(1):1–14, doi:10.1016/j.bbcan.2013.02.001 1
- [2] Watson, J., Crick, F. H. F., 1953, Molecular structure of nucleic acids, *Nature*, 171(4356):737–8, doi:10.1038/171737a0 1
- [3] DNA Model Crick-Watson by User:Alkivar - User:Alkivar. Licensed under Public domain via Wikimedia Commons 2
- [4] Template from Crick and Watson's DNA molecular model, 1953. (9660573227) by Science Museum London / Science and Society Picture Library - Template from Crick and Watson's DNA molecular model, 1953.Uploaded by Mrjohncummings. Licensed under Creative Comm 2
- [5] van Gunsteren, W. F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D. P., Glättli, A., Hünenberger, P. H., Kastenholz, M. a., Oostenbrink, C., Schenk, M., Trzesniak, D., van der Vegt, N. F. a., Yu, H. B., 2006, Biomolecular modeling: Goals, problems, perspectives., *Angewandte Chemie (International ed. in English)*, 45(25):4064–92, doi:10.1002/anie.200502655 1, 3
- [6] Leach, A. R., 2001, *Molecular Modelling: Principles and Applications*, Prentice Hall; Auflage: 2. A. 1, 10, 14, 28, 29
- [7] Mura, C., McAnany, C. E., 2014, An introduction to biomolecular simulations and docking, *Molecular Simulation*, 40(10-11):732–764, doi:10.1080/08927022.2014.935372 1, 12, 58
- [8] Zillig, W., Stetter, K. O., Wunderl, S., Schulz, W., Priess, H., Scholz, I., 1980, The Sulfolobus-?Caldariella? group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases, *Archives of Microbiology*, 125(3):259–269, doi:10.1007/BF00446886 4
- [9] She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P. M., Heikamp-de Jong, I., Jeffries, A. C., Kozera, C. J., Medina, N., Peng, X., Thi-Ngoc, H. P., Redder, P., Schenk, M. E., Theriault, C., Tolstrup, N., Charlebois, R. L., Doolittle, W. F., Duguet, M., Gaasterland, T., Garrett, R. A., Ragan, M. A., Sensen, C. W., Van der Oost, J., 2001, The complete genome of the crenarchaeon Sulfolobus solfataricus P2., *Proceedings of the National Academy of Sciences of the United States of America*, 98(14):7835–40, doi:10.1073/pnas.141222098 4
- [10] Wagner, M., Berkner, S., Ajon, M., Driessen, A. J. M., Lipps, G., Albers, S.-V., 2009, Expanding and understanding the genetic toolbox of the hyperthermophilic genus Sulfolobus., *Biochemical Society transactions*, 37(Pt 1):97–101, doi:10.1042/BST0370097 4

LITERATUR

- [11] Siebers, B., Schönheit, P., 2005, Unusual pathways and enzymes of central carbohydrate metabolism in Archaea., *Current opinion in microbiology*, 8(6):695–705, doi:10.1016/j.mib.2005.10.014 4
- [12] Snijders, A. P. L., Walther, J., Peter, S., Kinnman, I., de Vos, M. G. J., van de Werken, H. J. G., Brouns, S. J. J., van der Oost, J., Wright, P. C., 2006, Reconstruction of central carbon metabolism in *Sulfolobus solfataricus* using a two-dimensional gel electrophoresis map, stable isotope labelling and DNA microarray analysis., *Proteomics*, 6(5):1518–29, doi:10.1002/pmic.200402070 4
- [13] Albers, S.-V., Birkeland, N.-K. r., Driessen, A. J. M., Gertig, S., Haferkamp, P., Klenk, H.-P., Kouril, T., Manica, A., Pham, T. K., Ruoff, P., Schleper, C., Schomburg, D., Sharkey, K. J., Siebers, B., Sierocinski, P., Steuer, R., van der Oost, J., Westerhoff, H. V., Wieloch, P., Wright, P. C., Zaparty, M., 2009, SulfoSYS (*Sulfolobus* Systems Biology): towards a silicon cell model for the central carbohydrate metabolism of the archaeon *Sulfolobus solfataricus* under temperature variation., *Biochemical Society transactions*, 37(Pt 1):58–64, doi:10.1042/BST0370058 4
- [14] Ahmed, H., Ettema, T. J. G., Tjaden, B., Geerling, A. C. M., van der Oost, J., Siebers, B., 2005, The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation., *The Biochemical journal*, 390(Pt 2):529–40, doi:10.1042/BJ20041711 4
- [15] Lamble, H. J., Heyer, N. I., Bull, S. D., Hough, D. W., Danson, M. J., 2003, Metabolic pathway promiscuity in the archaeon *Sulfolobus solfataricus* revealed by studies on glucose dehydrogenase and 2-keto-3-deoxygluconate aldolase., *The Journal of biological chemistry*, 278(36):34066–72, doi:10.1074/jbc.M305818200 4, 6, 37, 44, 45
- [16] Lamble, H. J., Theodossis, A., Milburn, C. C., Taylor, G. L., Bull, S. D., Hough, D. W., Danson, M. J., 2005, Promiscuity in the part-phosphorylative Entner-Doudoroff pathway of the archaeon *Sulfolobus solfataricus*., *FEBS letters*, 579(30):6865–9, doi:10.1016/j.febslet.2005.11.028 4
- [17] Kim, S., Lee, S. B., 2005, Identification and characterization of *Sulfolobus solfataricus* D-gluconate dehydratase: a key enzyme in the non-phosphorylated Entner-Doudoroff pathway., *The Biochemical journal*, 387(Pt 1):271–80, doi:10.1042/BJ20041053 4
- [18] Kim, S., Lee, S. B., 2006, Characterization of *Sulfolobus solfataricus* 2-keto-3-deoxy-D-gluconate kinase in the modified Entner-Doudoroff pathway., *Bioscience, biotechnology, and biochemistry*, 70(6):1308–16, doi:10.1271/bbb.50566 4
- [19] Zaparty, M., Esser, D., Gertig, S., Haferkamp, P., Kouril, T., Manica, A., Pham, T. K., Reimann, J., Schreiber, K., Sierocinski, P., Teichmann, D., van Wolferen, M., von Jan, M., Wieloch, P., Albers, S. V., Driessen, A. J. M., Klenk, H.-P., Schleper, C., Schomburg, D., van der

LITERATUR

- Oost, J., Wright, P. C., Siebers, B., 2010, Hot standards for the thermoacidophilic archaeon *Sulfolobus solfataricus*., *Extremophiles : life under extreme conditions*, 14(1):119–42, doi:10.1007/s00792-009-0280-0 5
- [20] Becker, R. J., 1975, Report of the joint council of socio-economics of allergy., *Annals of allergy*, 35(2):109–12 6
- [21] Riveros-Rosas, H., Julián-Sánchez, A., Villalobos-Molina, R., Pardo, J. P., Piña, E., 2003, Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily., *European journal of biochemistry / FEBS*, 270(16):3309–34 6
- [22] Milburn, C. C., Lamb, H. J., Theodossis, A., Bull, S. D., Hough, D. W., Danson, M. J., Taylor, G. L., 2006, The structural basis of substrate promiscuity in glucose dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*., *The Journal of biological chemistry*, 281(21):14796–804, doi:10.1074/jbc.M601334200 6, 33, 34, 35, 36, 37, 38, 39
- [23] Eijssink, V. G. H., Vaaje-Kolstad, G., Vårum, K. M., Horn, S. J., 2008, Towards new enzymes for biofuels: lessons from chitinase research., *Trends in biotechnology*, 26(5):228–35, doi:10.1016/j.tibtech.2008.02.004 6, 7
- [24] Arakane, Y., Muthukrishnan, S., 2010, Insect chitinase and chitinase-like proteins., *Cellular and molecular life sciences : CMLS*, 67(2):201–16, doi:10.1007/s00018-009-0161-9 6, 7
- [25] Adrangi, S., Faramarzi, M. A., 2013, From bacteria to human: a journey into the world of chitinases., *Biotechnology advances*, 31(8):1786–95, doi:10.1016/j.biotechadv.2013.09.012 6
- [26] Kumar, A., Rao, M., 2010, A novel bifunctional peptidic aspartic protease inhibitor inhibits chitinase A from *Serratia marcescens*: Kinetic analysis of inhibition and binding affinity., *Biochimica et biophysica acta*, 1800(5):526–36, doi:10.1016/j.bbagen.2010.01.014 7
- [27] Pantoom, S., Vetter, I. R., Prinz, H., Suginta, W., 2011, Potent family-18 chitinase inhibitors: x-ray structures, affinities, and binding mechanisms., *The Journal of biological chemistry*, 286(27):24312–23, doi:10.1074/jbc.M110.183376 7
- [28] Volkmer, R., Tapia, V., Landgraf, C., 2012, Synthetic peptide arrays for investigating protein interaction domains., *FEBS letters*, 586(17):2780–6, doi:10.1016/j.febslet.2012.04.028 8
- [29] Kay, B. K., 2012, SH3 domains come of age., *FEBS letters*, 586(17):2606–8, doi:10.1016/j.febslet.2012.05.025 8
- [30] Panni, S., Dente, L., Cesareni, G., 2002, In vitro evolution of recognition specificity mediated by SH3 domains reveals target recognition rules., *The Journal of biological chemistry*, 277(24):21666–74, doi:10.1074/jbc.M109788200 8
- [31] Zafra-Ruano, A., Luque, I., 2012, Interfacial water molecules in SH3 interactions: Getting the full picture on polyproline recognition by protein-protein interaction domains., *FEBS letters*, 586(17):2619–30, doi:10.1016/j.febslet.2012.04.057 8, 76

LITERATUR

- [32] Chen, J. K., Schreiber, S. L., 1995, Combinatorial Synthesis and Multidimensional NMR Spectroscopy: An Approach to Understanding Protein–Ligand Interactions, *Angewandte Chemie International Edition in English*, 34(9):953–969, doi:10.1002/anie.199509531 9
- [33] Lim, W. A., Richards, F. M., Fox, R. O., 1994, Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains., *Nature*, 372(6504):375–9, doi:10.1038/372375a0 9
- [34] Hoffmann, S., Funke, S. A., Wiesehan, K., Moedder, S., Glück, J. M., Feuerstein, S., Gerdts, M., Mötter, J., Willbold, D., 2010, Competitively selected protein ligands pay their increase in specificity by a decrease in affinity., *Molecular bioSystems*, 6(1):126–33, doi:10.1039/b910945e 9
- [35] Musacchio, A., 2002, How SH3 domains recognize proline., *Advances in protein chemistry*, 61:211–68 9
- [36] Fernandez-Ballester, G., Blanes-Mira, C., Serrano, L., 2004, The tryptophan switch: changing ligand-binding specificity from type I to type II in SH3 domains., *Journal of molecular biology*, 335(2):619–29 9, 75
- [37] Schmidt, H., Hoffmann, S., Tran, T., Stoldt, M., Stangler, T., Wiesehan, K., Willbold, D., 2007, Solution structure of a Hck SH3 domain ligand complex reveals novel interaction modes., *Journal of molecular biology*, 365(5):1517–32, doi:10.1016/j.jmb.2006.11.013 9, 58, 59, 60, 65, 76, 77, 79, 81, 101, 103, 105, 106
- [38] Larson, S. M., Davidson, A. R., 2000, The identification of conserved interactions within the SH3 domain by alignment of sequences and structures., *Protein science : a publication of the Protein Society*, 9(11):2170–80, doi:10.1110/ps.9.11.2170 9, 60, 65, 75, 78, 98, 105
- [39] Trott, O., Olson, A. J., 2010, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading., *Journal of computational chemistry*, 31(2):455–61, doi:10.1002/jcc.21334 10, 49, 50
- [40] Zhang, X., Wong, S. E., Lightstone, F. C., 2013, Message passing interface and multithreading hybrid for parallel molecular docking of large databases on petascale high performance computing machines., *Journal of computational chemistry*, 34(11):915–27, doi:10.1002/jcc.23214 10
- [41] Zhang, X., Wong, S. E., Lightstone, F. C., 2014, Toward fully automated high performance computing drug discovery: a massively parallel virtual screening pipeline for docking and molecular mechanics/generalized born surface area rescoring to improve enrichment., *Journal of chemical information and modeling*, 54(1):324–37, doi:10.1021/ci4005145 10, 48
- [42] Blundell, T. L., Sibanda, B. L., Sternberg, M. J., Thornton, J. M., 1987, Knowledge-based prediction of protein structures and the design of novel molecules., *Nature*, 326(6111):347–52, doi:10.1038/326347a0 11

LITERATUR

- [43] Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., Sali, A., 2000, Comparative protein structure modeling of genes and genomes., *Annual review of biophysics and biomolecular structure*, 29:291–325, doi:10.1146/annurev.biophys.29.1.291 11
- [44] Sali, A., Blundell, T. L., 1993, Comparative protein modelling by satisfaction of spatial restraints., *Journal of molecular biology*, 234(3):779–815, doi:10.1006/jmbi.1993.1626 11
- [45] Zvelebil, M. J., Baum, J. O., 2008, *Understanding Bioinformatics*, Garland Science, 1st edition 11, 12
- [46] Berman, H. M., 2000, The Protein Data Bank, *Nucleic Acids Research*, 28(1):235–242, doi:10.1093/nar/28.1.235 11, 58
- [47] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., 1990, Basic local alignment search tool., *Journal of molecular biology*, 215(3):403–10, doi:10.1016/S0022-2836(05)80360-2 11
- [48] Pearson, W. R., Lipman, D. J., 1988, Improved tools for biological sequence comparison., *Proceedings of the National Academy of Sciences of the United States of America*, 85(8):2444–8 11
- [49] Sippl, M. J., 1993, Recognition of errors in three-dimensional structures of proteins., *Proteins*, 17(4):355–62, doi:10.1002/prot.340170404 12
- [50] Shen, M.-Y., Sali, A., 2006, Statistical potential for assessment and prediction of protein structures., *Protein science : a publication of the Protein Society*, 15(11):2507–24, doi:10.1110/ps.062416606 12
- [51] Blaney, J. M., Dixon, J. S., 1993, A good ligand is hard to find: Automated docking methods, *Perspectives in Drug Discovery and Design*, 1(2):301–319, doi:10.1007/BF02174531 12
- [52] Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., Ferrin, T. E., 1982, A geometric approach to macromolecule-ligand interactions, *Journal of Molecular Biology*, 161(2):269–288, doi:10.1016/0022-2836(82)90153-X 13
- [53] Goodsell, D. S., Olson, A. J., 1990, Automated docking of substrates to proteins by simulated annealing., *Proteins*, 8(3):195–202, doi:10.1002/prot.340080302 13
- [54] Judson, R., Jaeger, E., Treasurywala, A., 1994, A genetic algorithm based method for docking flexible molecules, *Journal of Molecular Structure: THEOCHEM*, 308:191–206, doi:10.1016/0166-1280(94)80102-9 13
- [55] Jones, G., Willett, P., Glen, R. C., 1995, Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation., *Journal of molecular biology*, 245(1):43–53 13

LITERATUR

- [56] Oshiro, C. M., Kuntz, I. D., Dixon, J. S., 1995, Flexible ligand docking using a genetic algorithm., *Journal of computer-aided molecular design*, 9(2):113–30 13
- [57] Klebe, G., 2009, *Wirkstoffdesign*, Spektrum Akademischer Verlag Heidelberg, 2nd edition 17, 20
- [58] Adcock, S. a., McCammon, J. A., 2006, Molecular dynamics: survey of methods for simulating the activity of proteins., *Chemical reviews*, 106(5):1589–615, doi:10.1021/cr040426m 18, 19, 20, 22, 25
- [59] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., Haak, J. R., 1984, Molecular dynamics with coupling to an external bath, *The Journal of Chemical Physics*, 81(8):3684, doi:10.1063/1.448118 18, 60
- [60] Nosé, S., 1984, A molecular dynamics method for simulations in the canonical ensemble, *Molecular Physics*, 52(2):255–268, doi:10.1080/00268978400101201 18
- [61] Hoover, W., 1985, Canonical dynamics: Equilibrium phase-space distributions, *Physical Review A*, 31(3):1695–1697, doi:10.1103/PhysRevA.31.1695 18
- [62] Ryckaert, J.-P., Ciccotti, G., Berendsen, H. J., 1977, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *Journal of Computational Physics*, 23(3):327–341, doi:10.1016/0021-9991(77)90098-5 18, 60
- [63] Hess, B., Bekker, H., Berendsen, H. J. C., Fraaije, J. G. E. M., 1997, LINCS: A linear constraint solver for molecular simulations, *Journal of Computational Chemistry*, 18(12):1463–1472, doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H 18
- [64] Miyamoto, S., Kollman, P. A., 1992, Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models, *Journal of Computational Chemistry*, 13(8):952–962, doi:10.1002/jcc.540130805 18
- [65] Darden, T., York, D., Pedersen, L., 1993, Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems, *The Journal of Chemical Physics*, 98(12):10089, doi:10.1063/1.464397 19
- [66] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., Pedersen, L. G., 1995, A smooth particle mesh Ewald method, *The Journal of Chemical Physics*, 103(19):8577, doi:10.1063/1.470117 19
- [67] Genheden, S., Ryde, U., 2011, A comparison of different initialization protocols to obtain statistically independent molecular dynamics simulations., *Journal of computational chemistry*, 32(2):187–95, doi:10.1002/jcc.21546 19, 80
- [68] Jorgensen, W. L., 1981, Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water, *Journal of the American Chemical Society*, 103(2):335–340, doi:10.1021/ja00392a016 20

LITERATUR

- [69] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Hermans, J., 1981, Interaction models for water in relation to protein hydration, *Intermolecular Forces*, pp. 331–342 20
- [70] Horn, H. W., Swope, W. C., Pitner, J. W., Madura, J. D., Dick, T. J., Hura, G. L., Head-Gordon, T., 2004, Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew., *The Journal of chemical physics*, 120(20):9665–78, doi:10.1063/1.1683075 20
- [71] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L., 1983, Comparison of simple potential functions for simulating liquid water, *The Journal of Chemical Physics*, 79(2):926, doi:10.1063/1.445869 20
- [72] Smith, P. E., Pettitt, B. M., 1994, Modeling Solvent in Biomolecular Systems, *The Journal of Physical Chemistry*, 98(39):9700–9711, doi:10.1021/j100090a002 20
- [73] Cisneros, G. A., Karttunen, M., Ren, P., Sagui, C., 2014, Classical electrostatics for biomolecular simulations., *Chemical reviews*, 114(1):779–814, doi:10.1021/cr300461d 23
- [74] Klamt, A., 2005, *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, Elsevier Science 23
- [75] Fogolari, F., Brigo, A., Molinari, H., 2002, The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology., *Journal of molecular recognition : JMR*, 15(6):377–92, doi:10.1002/jmr.577 23
- [76] Cisneros, G. A., Karttunen, M., Ren, P., Sagui, C., 2014, Classical electrostatics for biomolecular simulations., *Chemical reviews*, 114(1):779–814, doi:10.1021/cr300461d 23
- [77] Bashford, D., Case, D. A., 2000, Generalized born models of macromolecular solvation effects., *Annual review of physical chemistry*, 51:129–52, doi:10.1146/annurev.physchem.51.1.129 24
- [78] Hawkins, G. D., Cramer, C. J., Truhlar, D. G., 1996, Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium, *The Journal of Physical Chemistry*, 100(51):19824–19839, doi:10.1021/jp961710n 24, 61
- [79] Onufriev, A., Bashford, D., Case, D. a., 2004, Exploring protein native states and large-scale conformational changes with a modified generalized born model., *Proteins*, 55(2):383–94, doi:10.1002/prot.20033 24, 51
- [80] Mongan, J., Simmerling, C., McCammon, J. A., Case, D. A., Onufriev, A., 2007, Generalized Born model with a simple, robust molecular volume correction., *Journal of chemical theory and computation*, 3(1):156–169, doi:10.1021/ct600085e 24
- [81] Kongsted, J., Söderhjelm, P., Ryde, U., 2009, How accurate are continuum solvation models for drug-like molecules?, *Journal of computer-aided molecular design*, 23(7):395–409, doi:10.1007/s10822-009-9271-6 24

LITERATUR

- [82] Hermann, R. B., 1972, Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area, *The Journal of Physical Chemistry*, 76(19):2754–2759, doi:10.1021/j100663a023 24
- [83] Sitkoff, D., Sharp, K. A., Honig, B., 1994, Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models, *The Journal of Physical Chemistry*, 98(7):1978–1988, doi:10.1021/j100058a043 24
- [84] Tan, C., Tan, Y.-H., Luo, R., 2007, Implicit nonpolar solvent models., *The journal of physical chemistry. B*, 111(42):12263–74, doi:10.1021/jp073399n 24, 25
- [85] Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Kollman, P. A., 2010, *AMBER 11*, University of California, San Francisco 24, 60, 82
- [86] Marenich, A. V., Cramer, C. J., Truhlar, D. G., 2009, Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions., *The journal of physical chemistry. B*, 113(18):6378–96, doi:10.1021/jp810292n 25
- [87] Floris, F., Tomasi, J., 1989, Evaluation of the dispersion contribution to the solvation energy. A simple computational model in the continuum approximation, *Journal of Computational Chemistry*, 10(5):616–627, doi:10.1002/jcc.540100504 25
- [88] Barone, V., Cossi, M., Tomasi, J., 1997, A new definition of cavities for the computation of solvation free energies by the polarizable continuum model, *The Journal of Chemical Physics*, 107(8):3210, doi:10.1063/1.474671 25
- [89] Genheden, S., Mikulskis, P., Hu, L., Kongsted, J., Söderhjelm, P., Ryde, U., 2011, Accurate predictions of nonpolar solvation free energies require explicit consideration of binding-site hydration., *Journal of the American Chemical Society*, 133(33):13081–92, doi:10.1021/ja202972m 25
- [90] Genheden, S., Ryde, U., 2010, How to obtain statistically converged MM/GBSA results., *Journal of computational chemistry*, 31(4):837–46, doi:10.1002/jcc.21366 25, 60, 61
- [91] Jensen, F., 2011, *Introduction to Computational Chemistry*, WILEY-VCH Verlag GmbH & Co. KGaA, 2nd edition 26
- [92] Brooks, B., Karplus, M., 1983, Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor., *Proceedings of the National Academy of Sciences of the United States of America*, 80(21):6571–5 26

LITERATUR

- [93] Gourvil, J., Wechsler, J., Revuz, J., Touraine, R., 1979, [Eosinophilic fasciitis (Shulman syndrome) associated with insulin-dependent diabete. 1 case (author's transl)], *La Nouvelle presse médicale*, 8(50):4087–90 26
- [94] Levitt, M., Sander, C., Stern, P. S., 2009, The normal modes of a protein: Native bovine pancreatic trypsin inhibitor, *International Journal of Quantum Chemistry*, 24(S10):181–199, doi:10.1002/qua.560240721 26
- [95] Yang, L., Song, G., Jernigan, R. L., 2007, How well can we understand large-scale protein motions using normal modes of elastic network models?, *Biophysical journal*, 93(3):920–9, doi:10.1529/biophysj.106.095927 26
- [96] Bahar, I., Rader, A. J., 2005, Coarse-grained normal mode analysis in structural biology., *Current opinion in structural biology*, 15(5):586–92, doi:10.1016/j.sbi.2005.08.007 26
- [97] Li, M., Petukh, M., Alexov, E., Panchenko, A. R., 2014, Predicting the Impact of Missense Mutations on Protein-Protein Binding Affinity., *Journal of chemical theory and computation*, 10(4):1770–1780, doi:10.1021/ct401022c 26
- [98] Massova, I., Kollman, P. A., 1999, Computational Alanine Scanning To Probe Protein-Protein Interactions: A Novel Approach To Evaluate Binding Free Energies, *Journal of the American Chemical Society*, 121(36):8133–8143, doi:10.1021/ja990935j 26, 59
- [99] Huo, S., Massova, I., Kollman, P. A., 2002, Computational alanine scanning of the 1:1 human growth hormone-receptor complex., *Journal of computational chemistry*, 23(1):15–27, doi:10.1002/jcc.1153 26
- [100] Zhang, D., Nandi, S., Bryan, P., Pettit, S., Nguyen, D., Santos, M. A., Huang, N., 2010, Expression, purification, and characterization of recombinant human transferrin from rice (*Oryza sativa* L.), *Protein expression and purification*, 74(1):69–79, doi:10.1016/j.pep.2010.04.019 26
- [101] Moreira, I. S., Fernandes, P. A., Ramos, M. J., 2006, Unraveling the importance of protein-protein interaction: application of a computational alanine-scanning mutagenesis to the study of the IgG1 streptococcal protein G (C2 fragment) complex., *The journal of physical chemistry. B*, 110(22):10962–9, doi:10.1021/jp054760d 26
- [102] Gohlke, H., Kiel, C., Case, D. A., 2003, Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes., *Journal of molecular biology*, 330(4):891–913, doi:10.1016/S0022-2836(03)00610-7 26
- [103] Ferro, D. R., Hermans, J., 1977, A different best rigid-body molecular fit routine, *Acta Crystallographica Section A*, 33(2):345–347, doi:10.1107/S0567739477000862 27

LITERATUR

- [104] Kabsch, W., 1978, A discussion of the solution for the best rotation to relate two sets of vectors, *Acta Crystallographica Section A*, 34(5):827–828, doi:10.1107/S0567739478001680 27
- [105] Kuzmanic, A., Zagrovic, B., 2010, Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors., *Biophysical journal*, 98(5):861–71, doi:10.1016/j.bpj.2009.11.011 27
- [106] Shao, J., Tanner, S. W., Thompson, N., Cheatham, T. E., 2007, Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms, *Journal of Chemical Theory and Computation*, 3(6):2312–2334, doi:10.1021/ct700119m 28, 29, 83, 98
- [107] Jain, A. K., Murty, M. N., Flynn, P. J., 1999, Data clustering: a review, *ACM Computing Surveys*, 31(3):264–323, doi:10.1145/331499.331504 28
- [108] Torda, A. E., van Gunsteren, W. F., 1994, Algorithms for clustering molecular dynamics configurations, *Journal of Computational Chemistry*, 15(12):1331–1340, doi:10.1002/jcc.540151203 29
- [109] Jarvis, R., Patrick, E., 1973, Clustering Using a Similarity Measure Based on Shared Near Neighbors, *IEEE Transactions on Computers*, C-22(11):1025–1034, doi:10.1109/T-C.1973.223640 29
- [110] Downs, G. M., Willett, P., Fisanick, W., 1994, Similarity Searching and Clustering of Chemical-Structure Databases Using Molecular Property Data, *Journal of Chemical Information and Modeling*, 34(5):1094–1102, doi:10.1021/ci00021a011 30
- [111] Simmerling, C., Strockbine, B., Roitberg, A. E., 2002, All-Atom Structure Prediction and Folding Simulations of a Stable Protein, *Journal of the American Chemical Society*, 124(38):11258–11259, doi:10.1021/ja0273851 30
- [112] Mirjalili, V., Feig, M., 2013, Protein Structure Refinement through Structure Selection and Averaging from Molecular Dynamics Ensembles., *Journal of chemical theory and computation*, 9(2):1294–1303, doi:10.1021/ct300962x 30, 80
- [113] Knapp, B., Frantal, S., Cibena, M., Schreiner, W., Bauer, P., 2011, Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible?, *Journal of computational biology : a journal of computational molecular cell biology*, 18(8):997–1005, doi:10.1089/cmb.2010.0237 30, 80, 83, 107
- [114] Lyman, E., Zuckerman, D. M., 2006, Ensemble-based convergence analysis of biomolecular trajectories., *Biophysical journal*, 91(1):164–72, doi:10.1529/biophysj.106.082941 30, 80

LITERATUR

- [115] Genheden, S., Ryde, U., 2012, Will molecular dynamics simulations of proteins ever reach equilibrium?, *Physical chemistry chemical physics : PCCP*, pp. 8662–8677, doi:10.1039/c2cp23961b 30, 80
- [116] Lyman, E., Zuckerman, D. M., 2007, On the structural convergence of biomolecular simulations by determination of the effective sample size., *The journal of physical chemistry. B*, 111(44):12876–82, doi:10.1021/jp073061t 30, 80
- [117] Zuckerman, D. M., 2011, Equilibrium sampling in biomolecular simulations., *Annual review of biophysics*, 40:41–62, doi:10.1146/annurev-biophys-042910-155255 30, 80
- [118] Tusher, V. G., Tibshirani, R., Chu, G., 2001, Significance analysis of microarrays applied to the ionizing radiation response., *Proceedings of the National Academy of Sciences of the United States of America*, 98(9):5116–21, doi:10.1073/pnas.091062498 30, 59, 62, 80
- [119] Stanford's Robert Tibshirani on Significance Analysis of Microarrays - ScienceWatch.com 30
- [120] Significance Analysis of Microarrays 30
- [121] Schwender, H., 2011, siggenes: Multiple testing using SAM and Efron's empirical Bayes approaches 30, 62
- [122] Hastie, T., Tibshirani, R., Friedman, J., 2009, *The Elements of Statistical Learning - Data Mining, Inference, and Prediction, Second Edition*, Springer, 2nd edition 31
- [123] Kostrominova, T. Y., Dow, D. E., Dennis, R. G., Miller, R. A., Faulkner, J. A., 2005, Comparison of gene expression of 2-mo denervated, 2-mo stimulated-denervated, and control rat skeletal muscles., *Physiological genomics*, 22(2):227–43, doi:10.1152/physiolgenomics.00210.2004 32
- [124] Giardina, P., de Biasi, M. G., de Rosa, M., Gambacorta, A., Buonocore, V., 1986, Glucose dehydrogenase from the thermoacidophilic archaebacterium *Sulfolobus solfataricus*., *The Biochemical journal*, 239(3):517–22 33, 34
- [125] „Mutarotation D-Glucose V.1“ von Jü - Eigenes Werk. Lizenziert unter CC0 über Wikimedia Commons 33
- [126] McGuffin, L. J., Bryson, K., Jones, D. T., 2000, The PSIPRED protein structure prediction server., *Bioinformatics (Oxford, England)*, 16(4):404–5 34
- [127] Notredame, C., Higgins, D. G., Heringa, J., 2000, T-Coffee: A novel method for fast and accurate multiple sequence alignment., *Journal of molecular biology*, 302(1):205–17, doi:10.1006/jmbi.2000.4042 34
- [128] Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M.-Y., Pieper, U., Sali, A., 2007, Comparative protein structure modeling using MODELLER., *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]*, Chapter 2:Unit 2.9, doi:10.1002/0471140864.ps0209s50 34

LITERATUR

- [129] Schrödinger, L., 2009, The PyMOL Molecular Graphics System, Version 1.2r2 34, 35, 49, 50, 60
- [130] Baker, P. J., Britton, K. L., Fisher, M., Esclapez, J., Pire, C., Bonete, M. J., Ferrer, J., Rice, D. W., 2009, Active site dynamics in the zinc-dependent medium chain alcohol dehydrogenase superfamily., *Proceedings of the National Academy of Sciences of the United States of America*, 106(3):779–84, doi:10.1073/pnas.0807529106 34
- [131] Baker, N. A., Sept, D., Joseph, S., Holst, M. J., McCammon, J. A., 2001, Electrostatics of nanosystems: application to microtubules and the ribosome., *Proceedings of the National Academy of Sciences of the United States of America*, 98(18):10037–41, doi:10.1073/pnas.181342398 35
- [132] Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., Baker, N. A., 2004, PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations., *Nucleic acids research*, 32(Web Server issue):W665–7, doi:10.1093/nar/gkh381 35
- [133] Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., Baker, N. A., 2007, PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations., *Nucleic acids research*, 35(Web Server issue):W522–5, doi:10.1093/nar/gkm276 35
- [134] Li, H., Robertson, A. D., Jensen, J. H., 2005, Very fast empirical prediction and rationalization of protein pKa values., *Proteins*, 61(4):704–21, doi:10.1002/prot.20660 35
- [135] Dybowski, J. N., 2006, *Development of a method for optimal superposition of pairs of similar macromolecules*, Diploma thesis, University of Applied Sciences, Bingen 35
- [136] Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., Olson, A. J., 1998, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *Journal of Computational Chemistry*, 19(14):1639–1662, doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B 35
- [137] Krieger, E., Koraimann, G., Vriend, G., 2002, Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field., *Proteins*, 47(3):393–402 35
- [138] Petter, C., Scholz, C., Wessner, H., Hansen, G., Henklein, P., Watanabe, T., Höhne, W., 2008, Phage display screening for peptidic chitinase inhibitors., *Journal of molecular recognition : JMR*, 21(6):401–9, doi:10.1002/jmr.911 46
- [139] Sewczyk, K., 2010, *Berechnungen von Protein-Ligand-Wechselwirkungen am Beispiel spezieller Mutasen und Chitinasen*, Ph.D. thesis, Universität Duisburg-Essen 47
- [140] Yang, C.-Y., Sun, H., Chen, J., Nikolovska-Coleska, Z., Wang, S., 2009, Importance of ligand reorganization free energy in protein-ligand binding-affinity prediction., *Journal of the American Chemical Society*, 131(38):13709–21, doi:10.1021/ja9039373 47

LITERATUR

- [141] Gouda, H., Sunazuka, T., Iguchi, K., Sugawara, A., Hirose, T., Noguchi, Y., Saito, Y., Yanaï, Y., Yamamoto, T., Watanabe, T., Shiomi, K., Omura, S., Hirono, S., 2009, Computer-aided rational molecular design of argifin-derivatives with increased inhibitory activity against chitinase B from *Serratia marcescens*., *Bioorganic & medicinal chemistry letters*, 19(10):2630–3, doi:10.1016/j.bmcl.2009.04.013 47
- [142] Gouda, H., Sunazuka, T., Hirose, T., Iguchi, K., Yamaotsu, N., Sugawara, A., Noguchi, Y., Saito, Y., Yamamoto, T., Watanabe, T., Shiomi, K., Omura, S., Hirono, S., 2010, NMR spectroscopy and computational analysis of interaction between *Serratia marcescens* chitinase B and a dipeptide derived from natural-product cyclopentapeptide chitinase inhibitor argifin., *Bioorganic & medicinal chemistry*, 18(16):5835–44, doi:10.1016/j.bmc.2010.06.093 47, 48, 56, 57
- [143] Antunes, D. A., Vieira, G. F., Rigo, M. M., Cibulski, S. P., Sinigaglia, M., Chies, J. A. B., 2010, Structural allele-specific patterns adopted by epitopes in the MHC-I cleft and reconstruction of MHC:peptide complexes to cross-reactivity assessment., *PloS one*, 5(4):e10353, doi:10.1371/journal.pone.0010353 47, 57
- [144] Azoia, N. G., Fernandes, M. M., Micaêlo, N. M., Soares, C. M., Cavaco-Paulo, A., 2012, Molecular modeling of hair keratin/peptide complex: Using MM-PBSA calculations to describe experimental binding results., *Proteins*, 80(5):1409–17, doi:10.1002/prot.24037 48
- [145] Fu, C.-w., Wang, Y.-P., Fang, T.-Y., Lin, T.-H., 2013, Interaction between trehalose and MTHase from *Sulfolobus solfataricus* studied by theoretical computation and site-directed mutagenesis., *PloS one*, 8(7):e68565, doi:10.1371/journal.pone.0068565 48
- [146] Hou, T., Wang, J., Li, Y., Wang, W., 2011, Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking., *Journal of computational chemistry*, 32(5):866–77, doi:10.1002/jcc.21666 48, 55, 57
- [147] Hou, T., Wang, J., Li, Y., Wang, W., 2011, Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations., *Journal of chemical information and modeling*, 51(1):69–82, doi:10.1021/ci100275a 48
- [148] Kokubo, H., Tanaka, T., Okamoto, Y., 2013, Prediction of Protein–Ligand Binding Structures by Replica-Exchange Umbrella Sampling Simulations: Application to Kinase Systems, *Journal of Chemical Theory and Computation*, 9(10):4660–4671, doi:10.1021/ct4004383 48
- [149] Mikulskis, P., Genheden, S., Rydberg, P., Sandberg, L., Olsen, L., Ryde, U., 2012, Binding affinities in the SAMPL3 trypsin and host-guest blind tests estimated with the MM/PBSA and LIE methods., *Journal of computer-aided molecular design*, 26(5):527–41, doi:10.1007/s10822-011-9524-z 48

LITERATUR

- [150] Rastelli, G., Degliesposti, G., Del Rio, A., Sgobba, M., 2009, Binding estimation after refinement, a new automated procedure for the refinement and rescoring of docked ligands in virtual screening., *Chemical biology & drug design*, 73(3):283–6, doi:10.1111/j.1747-0285.2009.00780.x 48, 55
- [151] Sgobba, M., Caporuscio, F., Anighoro, A., Portioli, C., Rastelli, G., 2012, Application of a post-docking procedure based on MM-PBSA and MM-GBSA on single and multiple protein conformations., *European journal of medicinal chemistry*, 58:431–40, doi:10.1016/j.ejmech.2012.10.024 48
- [152] Suenaga, A., Okimoto, N., Hirano, Y., Fukui, K., 2012, An efficient computational method for calculating ligand binding affinities., *PloS one*, 7(8):e42846, doi:10.1371/journal.pone.0042846 48
- [153] Sun, H., Li, Y., Shen, M., Tian, S., Xu, L., Pan, P., Guan, Y., Hou, T., 2014, Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring., *Physical chemistry chemical physics : PCCP*, 16(40):22035–45, doi:10.1039/c4cp03179b 48
- [154] Zhu, T., Lee, H., Lei, H., Jones, C., Patel, K., Johnson, M. E., Hevener, K. E., 2013, Fragment-based drug discovery using a multidomain, parallel MD-MM/PBSA screening protocol., *Journal of chemical information and modeling*, 53(3):560–72, doi:10.1021/ci300502h 48
- [155] Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuik, Y., McGinnis, S., Madden, T. L., 2008, NCBI BLAST: a better web interface., *Nucleic acids research*, 36(Web Server issue):W5–9, doi:10.1093/nar/gkn201 49
- [156] Craig, I. R., Pflieger, C., Gohlke, H., Essex, J. W., Spiegel, K., 2011, Pocket-space maps to identify novel binding-site conformations in proteins., *Journal of chemical information and modeling*, 51(10):2666–79, doi:10.1021/ci200168b 50
- [157] van Aalten, D. M., Komander, D., Synstad, B., Gå seidnes, S., Peter, M. G., Eijssink, V. G., 2001, Structural insights into the catalytic mechanism of a family 18 exo-chitinase., *Proceedings of the National Academy of Sciences of the United States of America*, 98(16):8979–84, doi:10.1073/pnas.151103798 50
- [158] Vaaje-Kolstad, G., Houston, D. R., Rao, F. V., Peter, M. G., Synstad, B. r., van Aalten, D. M. F., Eijssink, V. G. H., 2004, Structure of the D142N mutant of the family 18 chitinase ChiB from *Serratia marcescens* and its complex with allosamidin., *Biochimica et biophysica acta*, 1696(1):103–11 50
- [159] Houston, D. R., Synstad, B. r., Eijssink, V. G. H., Stark, M. J. R., Eggleston, I. M., van Aalten, D. M. F., 2004, Structure-based exploration of cyclic dipeptide chitinase inhibitors., *Journal of medicinal chemistry*, 47(23):5713–20, doi:10.1021/jm049940a 50

LITERATUR

- [160] Vaaje-Kolstad, G., Vasella, A., Peter, M. G., Netter, C., Houston, D. R., Westereng, B. r., Synstad, B. r., Eijsink, V. G. H., van Aalten, D. M. F., 2004, Interactions of a family 18 chitinase with the designed inhibitor HM508 and its degradation product, chitobiono-delta-lactone., *The Journal of biological chemistry*, 279(5):3612–9, doi:10.1074/jbc.M310057200 50
- [161] Houston, D. R., Eggleston, I., Synstad, B. r., Eijsink, V. G. H., van Aalten, D. M. F., 2002, The cyclic dipeptide CI-4 [cyclo-(l-Arg-d-Pro)] inhibits family 18 chitinases by structural mimicry of a reaction intermediate., *The Biochemical journal*, 368(Pt 1):23–7, doi:10.1042/BJ20021034 50
- [162] Tabudravu, J. N., Eijsink, V. G. H., Gooday, G. W., Jaspars, M., Komander, D., Legg, M., Synstad, B., van Aalten, D. M. F., 2002, Psammaplin A, a chitinase inhibitor isolated from the Fijian marine sponge *Aplysinella rhax*., *Bioorganic & medicinal chemistry*, 10(4):1123–8 50
- [163] Kolstad, G., Synstad, B., Eijsink, V. G. H., van Aalten, D. M. F., 2002, Structure of the D140N mutant of chitinase B from *Serratia marcescens* at 1.45 Å resolution., *Acta crystallographica. Section D, Biological crystallography*, 58(Pt 2):377–9, doi:10.1107/S0907444901018972 50
- [164] van Aalten, D. M., Synstad, B., Brurberg, M. B., Hough, E., Riise, B. W., Eijsink, V. G., Wierenga, R. K., 2000, Structure of a two-domain chitotriosidase from *Serratia marcescens* at 1.9-Å resolution., *Proceedings of the National Academy of Sciences of the United States of America*, 97(11):5842–7 50
- [165] Berendsen, H., van der Spoel, D., van Drunen, R., 1995, GROMACS: A message-passing parallel molecular dynamics implementation, *Computer Physics Communications*, 91(1-3):43–56, doi:10.1016/0010-4655(95)00042-E 49
- [166] Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., Kollman, P., 2003, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations., *Journal of computational chemistry*, 24(16):1999–2012, doi:10.1002/jcc.10349 51
- [167] Miller, B. R., McGee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., Roitberg, A. E., 2012, MMPBSA.py : An Efficient Program for End-State Free Energy Calculations, *Journal of Chemical Theory and Computation*, 8(9):3314–3321, doi:10.1021/ct300418h 51, 61
- [168] Wang, R., Wang, S., 2001, How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment, *Journal of Chemical Information and Modeling*, 41(5):1422–1426, doi:10.1021/ci010025x 56

LITERATUR

- [169] Oda, A., Tsuchida, K., Takakura, T., Yamaotsu, N., Hirono, S., 2005, Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes., *Journal of chemical information and modeling*, 46(1):380–91, doi:10.1021/ci050283k 56
- [170] McCammon, J. A., Gelin, B. R., Karplus, M., 1977, Dynamics of folded proteins, *Nature*, 267(5612):585–590, doi:10.1038/267585a0 58
- [171] Papaleo, E., Pasi, M., Tiberti, M., De Gioia, L., 2011, Molecular dynamics of mesophilic-like mutants of a cold-adapted enzyme: insights into distal effects induced by the mutations., *PloS one*, 6(9):e24214, doi:10.1371/journal.pone.0024214 59
- [172] Piana, S., Lindorff-Larsen, K., Dirks, R. M., Salmon, J. K., Dror, R. O., Shaw, D. E., 2012, Evaluating the effects of cutoffs and treatment of long-range electrostatics in protein folding simulations., *PloS one*, 7(6):e39918, doi:10.1371/journal.pone.0039918 60, 82
- [173] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L., 1983, Comparison of simple potential functions for simulating liquid water, *The Journal of Chemical Physics*, 79(2):926, doi:10.1063/1.445869 60, 82
- [174] Ewald, P. P., 1921, Die Berechnung optischer und elektrostatischer Gitterpotentiale, *Annalen der Physik*, 369(3):253–287, doi:10.1002/andp.19213690304 61
- [175] Hawkins, G. D., Cramer, C. J., Truhlar, D. G., 1995, Pairwise solute descreening of solute charges from a dielectric medium, *Chemical Physics Letters*, 246(1-2):122–129, doi:10.1016/0009-2614(95)01082-K 61
- [176] R Development Core Team, 2010, R: A Language and Environment for Statistical Computing 61, 63
- [177] Roe, D. R., Cheatham, T. E., 2013, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *Journal of Chemical Theory and Computation*, 9(7):3084–3095, doi:10.1021/ct400341p 61, 82
- [178] Durrant, J. D., McCammon, J. A., 2011, Molecular dynamics simulations and drug discovery., *BMC biology*, 9:71, doi:10.1186/1741-7007-9-71 65
- [179] van Aalten, D. M., Amadei, A., Bywater, R., Findlay, J. B., Berendsen, H. J., Sander, C., Stouten, P. F., 1996, A comparison of structural and dynamic properties of different simulation methods applied to SH3., *Biophysical journal*, 70(2):684–92, doi:10.1016/S0006-3495(96)79608-X 65
- [180] Arold, S., O’Brien, R., Franken, P., Strub, M. P., Hoh, F., Dumas, C., Ladbury, J. E., 1998, RT loop flexibility enhances the specificity of Src family SH3 domains for HIV-1 Nef., *Biochemistry*, 37(42):14683–91, doi:10.1021/bi980989q 65
- [181] Hiipakka, M., Saksela, K., 2007, Versatile retargeting of SH3 domain binding by modification of non-conserved loop residues., *FEBS letters*, 581(9):1735–41, doi:10.1016/j.febslet.2007.03.044 65

LITERATUR

- [182] Pene-Dumitrescu, T., Shu, S. T., Wales, T. E., Alvarado, J. J., Shi, H., Narute, P., Moroco, J. a., Yeh, J. I., Engen, J. R., Smithgall, T. E., 2012, HIV-1 Nef interaction influences the ATP-binding site of the Src-family kinase, Hck., *BMC chemical biology*, 12(1):1, doi:10.1186/1472-6769-12-1 65
- [183] Shi, X., Betzi, S., Lugari, A., Opi, S., Restouin, A., Parrot, I., Martinez, J., Zimmermann, P., Lecine, P., Huang, M., Arold, S. T., Collette, Y., Morelli, X., 2012, Structural recognition mechanisms between human Src homology domain 3 (SH3) and ALG-2-interacting protein X (Alix)., *FEBS letters*, 586(13):1759–64, doi:10.1016/j.febslet.2012.05.017 65
- [184] Tran, T., Hoffmann, S., Wiesehan, K., Jonas, E., Luge, C., Aladag, A., Willbold, D., 2005, Insights into human Lck SH3 domain binding specificity: different binding modes of artificial and native ligands., *Biochemistry*, 44(45):15042–52, doi:10.1021/bi051403k 76, 79, 108
- [185] Batra-Safferling, R., Granzin, J., Mödder, S., Hoffmann, S., Willbold, D., 2010, Structural studies of the phosphatidylinositol 3-kinase (PI3K) SH3 domain in complex with a peptide ligand: role of the anchor residue in ligand binding., *Biological chemistry*, 391(1):33–42, doi:10.1515/BC.2010.003 76, 77, 79, 81, 101, 106, 108
- [186] Matsumura, Y., Shinjo, M., Kim, S. J., Okishio, N., Gruebele, M., Kihara, H., 2013, Transient helical structure during PI3K and Fyn SH3 domain folding., *The journal of physical chemistry. B*, 117(17):4836–43, doi:10.1021/jp400167s 79
- [187] Dasgupta, A., Udgaonkar, J. B., 2012, Four-state folding of a SH3 domain: salt-induced modulation of the stabilities of the intermediates and native state., *Biochemistry*, 51(23):4723–34, doi:10.1021/bi300223b 79
- [188] Ahmad, M., Gu, W., Helms, V., 2008, Mechanism of fast peptide recognition by SH3 domains., *Angewandte Chemie (International ed. in English)*, 47(40):7626–30, doi:10.1002/anie.200801856 79
- [189] Buch, I., Giorgino, T., De Fabritiis, G., 2011, Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations., *Proceedings of the National Academy of Sciences of the United States of America*, 108(25):10184–9, doi:10.1073/pnas.1103547108 79
- [190] Shan, Y., Kim, E. T., Eastwood, M. P., Dror, R. O., Seeliger, M. a., Shaw, D. E., 2011, How does a drug molecule find its target binding site?, *Journal of the American Chemical Society*, 133(24):9181–3, doi:10.1021/ja202726y 79
- [191] Cao, Y., Song, L., Miao, Z., Hu, Y., Tian, L., Jiang, T., 2011, Improved side-chain modeling by coupling clash-detection guided iterative search with rotamer relaxation., *Bioinformatics (Oxford, England)*, 27(6):785–90, doi:10.1093/bioinformatics/btr009 80

LITERATUR

- [192] Lopes, A., Alexandrov, A., Bathelt, C., Archontis, G., Simonson, T., 2007, Computational sidechain placement and protein mutagenesis with implicit solvent models., *Proteins*, 67(4):853–67, doi:10.1002/prot.21379 80
- [193] Eyal, E., Najmanovich, R., Edelman, M., Sobolev, V., 2003, Protein side-chain rearrangement in regions of point mutations., *Proteins*, 50(2):272–82, doi:10.1002/prot.10276 80
- [194] Zeiske, T., Stafford, K. a., Friesner, R. a., Palmer, A. G., 2013, Starting-structure dependence of nanosecond timescale intersubstate transitions and reproducibility of MD-derived order parameters., *Proteins*, 81(3):499–509, doi:10.1002/prot.24209 80
- [195] Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A. W., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Hsieh, M. J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., Kollman, P. A., 2012, *AMBER 12*, University of California, San Francisco 82
- [196] Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M., Jensen, J. H., 2011, PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions, *Journal of Chemical Theory and Computation*, 7(2):525–537, doi:10.1021/ct100578z 82
- [197] Williams, J. C., Wierenga, R. K., Saraste, M., 1998, Insights into Src kinase functions: structural comparisons, *Trends in Biochemical Sciences*, 23(5):179–184, doi:10.1016/S0968-0004(98)01202-X 103
- [198] Kuhn, O., 2013, *Binding Free Energy Calculations and Molecular Dynamics Studies on Complexes of Viral Proteases with their Ligands*, Ph.D. thesis, Universität Duisburg-Essen, Germany 107
- [199] Pawson, T., Nash, P., 2003, Assembly of cell regulatory systems through protein interaction domains., *Science (New York, N.Y.)*, 300(5618):445–52, doi:10.1126/science.1083653 107
- [200] Obsil, T., Obsilova, V., 2011, Structural basis of 14-3-3 protein functions., *Seminars in cell & developmental biology*, 22(7):663–72, doi:10.1016/j.semcdb.2011.09.001 107
- [201] Molzan, M., Schumacher, B., Ottmann, C., Baljuls, A., Polzien, L., Weyand, M., Thiel, P., Rose, R., Rose, M., Kuhenne, P., Kaiser, M., Rapp, U. R., Kuhlmann, J., Ottmann, C., 2010, Impaired binding of 14-3-3 to C-RAF in Noonan syndrome suggests new approaches in diseases with increased Ras signaling., *Molecular and cellular biology*, 30(19):4698–711, doi:10.1128/MCB.01636-09 107
- [202] Thiel, P., Röglin, L., Meissner, N., Hennig, S., Kohlbacher, O., Ottmann, C., 2013, Virtual screening and experimental validation reveal novel small-molecule inhibitors of 14-3-3 protein-protein interactions., *Chemical communications (Cambridge, England)*, 49(76):8468–70, doi:10.1039/c3cc44612c 107

LITERATUR

- [203] Rittinger, K., Budman, J., Xu, J., Volinia, S., Cantley, L. C., Smerdon, S. J., Gamblin, S. J., Yaffe, M. B., 1999, Structural analysis of 14-3-3 phosphopeptide complexes identifies a dual role for the nuclear export signal of 14-3-3 in ligand binding., *Molecular cell*, 4(2):153–66 107

Publikation

Patrick Haferkamp, Simone Kutschki, Jenny Treichel, Hatim Hemeda, **Karsten Sewczyk**, Daniel Hoffmann, Melanie Zaparty, and Bettina Siebers. An additional glucose dehydrogenase from *Sulfolobus solfataricus*: fine-tuning of sugar degradation? *Biochemical Society transactions*, 2011, 39(1):77-81.

An additional glucose dehydrogenase from *Sulfolobus solfataricus*: fine-tuning of sugar degradation?

Patrick Haferkamp*, Simone Kutschki*, Jenny Treichel*, Hatim Hemeda†, Karsten Sewczyk‡, Daniel Hoffmann‡, Melanie Zaparty*¹ and Bettina Siebers*¹

*Faculty of Chemistry, Biofilm Centre, Molecular Enzyme Technology and Biochemistry, University of Duisburg-Essen, Universitätsstrasse, Building S03 V03 F44, 45117 Essen, Germany, †Department of Otorhinolaryngology, University Hospital of Essen, Hufelandstrasse 55, 45147 Essen, Germany, and ‡Faculty of Biology and Geography, Center for Medical Biotechnology, Bioinformatics, University of Duisburg-Essen, Universitätsstrasse, Building S03 S03 A26, 45117 Essen, Germany

Abstract

Within the SulfoSYS (*Sulfolobus* Systems Biology) project, the effect of temperature on a metabolic network is investigated at the systems level. *Sulfolobus solfataricus* utilizes an unusual branched ED (Entner–Doudoroff) pathway for sugar degradation that is promiscuous for glucose and galactose. In the course of metabolic pathway reconstruction, a glucose dehydrogenase isoenzyme (GDH-2, SS03204) was identified. GDH-2 exhibits high similarity to the previously characterized GDH-1 (SS03003, 61 % amino acid identity), but possesses different enzymatic properties, particularly regarding substrate specificity and catalytic efficiency. In contrast with GDH-1, which exhibits broad substrate specificity for C₅ and C₆ sugars, GDH-2 is absolutely specific for glucose. The comparison of kinetic parameters suggests that GDH-2 might represent the major player in glucose catabolism via the branched ED pathway, whereas GDH-1 might have a dominant role in galactose degradation via the same pathway as well as in different sugar-degradation pathways.

Introduction

The thermoacidophilic archaeon *Sulfolobus solfataricus* (strain P2, DSM 1617) was originally isolated from a solfataric field in Pisciarelli (Italy). The organism optimally grows at 80°C (range 60–92°C) and pH 3.5 (range 2–4) [1]. *S. solfataricus* has been established as model organism within the (hyper)thermophilic archaea. The complete genome sequence is available [2], genetic systems have been established [3], numerous biochemical studies have been performed (for a review, see [4]), and functional genomics approaches, including transcriptomics, proteomics, comparative genomics, as well as a systems biology approach have been applied [5–7].

In course of the SulfoSYS (*Sulfolobus* Systems Biology) [7] project, the CCM (central carbohydrate metabolism) and the effect of temperature change on a metabolic network at systems level is under investigation by integrating data from the different ‘-omics’ techniques (genomics, transcriptomics, proteomics and metabolomics) as well as enzymatic information to build a silicon cell model [4,6]. An essential prerequisite of the project is the

reconstruction and identification of players in the CCM network of *S. solfataricus* in order to provide detailed enzymatic/biochemical information for enzymes and to study the behaviour and regulation of the network under temperature change.

Like other (hyper)thermophilic archaea (e.g. *Thermoproteus tenax* [8]), *S. solfataricus* uses an unusual branched ED (Entner–Doudoroff) pathway for sugar degradation that is characterized by an np (non-phosphorylative) and an sp (semi-phosphorylative) branch [8,9]. In contrast with the classical ED pathway, glucose is not phosphorylated at the beginning of the pathway, but is converted via gluconate into KDG (2-keto-3-desoxygluconate), the characteristic intermediate of the pathway. KDG is either phosphorylated by KDG kinase forming KDPG (2-keto-3-desoxy-6-phosphogluconate) (sp branch) or cleaved directly by the bifunctional KD(P)G [2-keto-3-desoxy-(6-phospho)-gluconate] aldolase and phosphorylation takes place at the level of glycerate (np branch).

Interestingly, the pathway has been shown to be promiscuous for glucose and galactose degradation in *S. solfataricus* [9]: GDH (glucose dehydrogenase) (EC 1.1.1.47), gluconate dehydratase, KDG kinase and KD(P)G aldolase catalyse the degradation of both substrates [9–12].

The first enzyme of the pathway, GDH, catalyses the NAD(P)⁺-dependent oxidation of glucose to gluconate. The enzyme has been purified previously from cell extracts of *S. solfataricus* and has been shown to possess dual

Key words: central carbohydrate metabolism, crenarchaeon, Entner–Doudoroff pathway, glucose dehydrogenase, glycolysis, *Sulfolobus solfataricus*, *Sulfolobus* Systems Biology (SulfoSYS).

Abbreviations used: CCM, central carbohydrate metabolism; ED, Entner–Doudoroff; GDH, glucose dehydrogenase; KDG, 2-keto-3-desoxygluconate; KD(P)G, 2-keto-3-desoxy-(6-phospho)-gluconate; np, non-phosphorylative; sp, semi-phosphorylative; SulfoSYS, *Sulfolobus* Systems Biology.

¹ To whom correspondence should be addressed (email bettina.siebers@uni-due.de).

co-substrate ($\text{NAD}^+/\text{NADP}^+$) and broad substrate specificity [13]. Detailed enzymatic characterization of the recombinant enzyme (SSO3003 [9]) and the available crystal structure [14] allowed for further insights into the catalytic activity and promiscuity of GDH.

GDH belongs to the MDR (medium-chain alcohol/polyol dehydrogenase/reductase) branch of the pyridine nucleotide-dependent alcohol/polyol/sugar dehydrogenase superfamily [15]. In the course of pathway reconstruction (M. Zaparty, T. Kouril, D. Esser, P. Haferkamp, J. Reimann, S.V. Albers and B. Siebers, unpublished work), several members of this superfamily were identified, raising questions about possible GDH isoenzymes in *S. solfataricus*. In previous studies, three alcohol dehydrogenases have been characterized (SSO2536, SSO1646 and SSO0764) [16]. However, since it is difficult to predict function by sequence alone, the additional GDH candidate (SSO3204) was analysed. The protein shows high similarity (61 % amino acid sequence identity) to the previously characterized GDH of *S. solfataricus* (SSO3003) [9,14].

We have therefore cloned the gene encoding the putative GDH (SSO3204), expressed the protein heterologously in *Escherichia coli*, purified it using ion-exchange chromatography and gel filtration, and analysed (co-)substrate specificity and its physiological role in the CCM of *S. solfataricus*.

Enzymatic characterization of GDH-2 and comparison with GDH-1

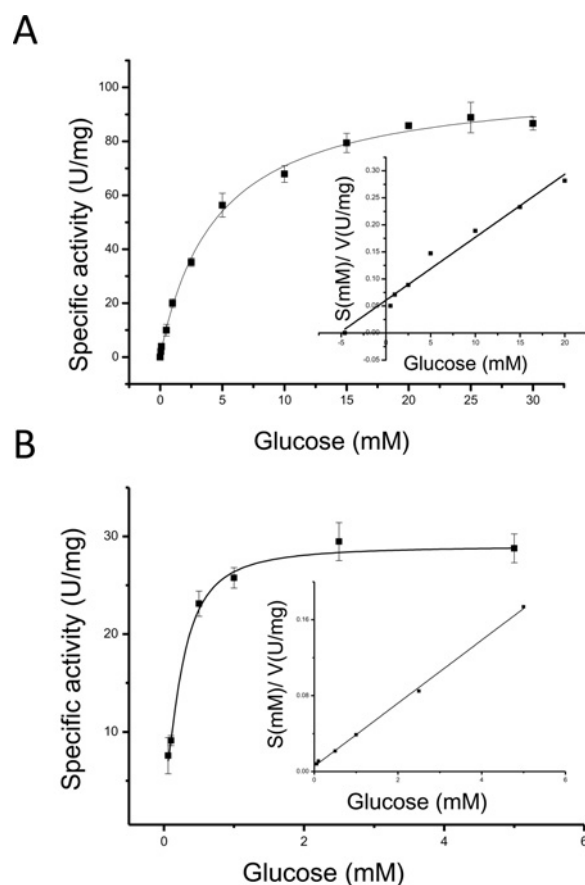
The putative GDH gene (SSO3204) was cloned and the recombinant protein was expressed and purified. From 6.2 g of cells (wet weight) 1.52 mg of enzyme with a specific GDH activity of 97 units/mg protein was recovered [17].

Like the previously characterized GDH-1 (SSO3003), GDH-2 (SSO3204) catalyses the irreversible oxidation of D-glucose forming gluconate with dual co-substrate ($\text{NAD}^+/\text{NADP}^+$) specificity. Remarkably, in contrast with GDH-1, which catalyses the oxidation of more than ten different C_5 and C_6 sugars [14], GDH-2 is absolutely specific for D-glucose; D-galactose, D-allose, D-mannose, D-xylose, L-arabinose, D-ribose, D-xylose, D-arabinose, L-xylose, D-glucosamine, 2-deoxy-D-glucose, D-fucose, D-lactose, D-maltose, D-fructose and ethanol were not used as substrates.

Interestingly, the oxidation rate of the different sugars by GDH-1 depends significantly upon the co-substrate available [9,13] [e.g. relative rates compared with D-glucose (100 % NADP^+ , 83 % NAD^+); L-arabinose, (99 %, NAD^+ , 95 % NADP^+); D-galactose (53 %, NAD^+ , 12 %, NADP^+); D-xylose (72 %, NAD^+ , 8 %, NADP^+)] [9]. GDH-2 follows classical Michaelis–Menten kinetics for glucose with NAD^+ and NADP^+ (10 mM) (Figure 1, Table 1). Strikingly, for both co-substrates, a significant inhibition has been observed at higher (>12 mM) concentrations, which is currently under investigation. For both enzymes, higher activity (GDH-1: 1.6-fold; GDH-2: 3.5-fold) was observed in the presence of NAD^+ at 70°C. However, in the presence of NADP^+ ,

Figure 1 | GDH-2 activity assay

GDH-2 (SSO3204) activity assayed at 70°C with either NAD^+ (A) or NADP^+ (B) as co-substrate (10 mM). The assay was performed in 100 mM Hepes (pH 6.5) in the presence of 10 mM MgCl_2 . Insets show the respective Hanes–Woolf plot. All measurements were performed in triplicate and are means \pm S.D. Kinetic parameters are given in Table 1.



GDH-2 exhibits a significantly (27-fold) higher affinity towards glucose [0.17 mM (NADP^+); 4.59 mM (NAD^+)], whereas GDH-1 shows no difference with respect to co-substrates [1.3 mM (NADP^+); 1.5 mM (NAD^+)]. Therefore, in comparison with GDH-1, the catalytic efficiency in the presence of NADP^+ is significantly increased (k_{cat}/K_m for GDH-2 and GDH-1 of 115.17 and 36.7 $\text{s}^{-1} \cdot \text{mM}^{-1}$ respectively), suggesting that GDH-2 is the preferred enzyme in glucose catabolism. However, GDH-1 seems to be slightly preferred in the presence of NAD^+ (k_{cat}/K_m for GDH-2 and GDH-1 of 49.9 and 14.86 $\text{s}^{-1} \cdot \text{mM}^{-1}$ respectively) and at high glucose concentration (Table 1).

Structural and model comparison

Beside kinetic parameters, the major difference between both isoenzymes is their substrate specificity. As mentioned above, both enzymes share 61 % amino acid sequence identity and an amino acid sequence alignment reveals the presence of the conserved structural zinc-binding sites (four conserved cysteine residues, Cys⁹³, Cys⁹⁶, Cys⁹⁹ and Cys¹⁰⁷ in GDH-1)

Table 1 | Comparison of kinetic parameters of GDH-1 (SSO3003) [9,14,22,23] and GDH-2 (SSO3204) from *S. solfataricus* determined at 70°C

Substrate	GDH-1 (SSO3003)		GDH-2 (S03204)	
D-Glucose				
Co-substrate	NAD ⁺	NADP ⁺	NAD ⁺	NADP ⁺
K_m (mM)	1.5	1.3	4.59	0.17
V_{max} (units/mg)	110	70	102.42	29.03
k_{cat} (s ⁻¹)	74.9	47.7	68.23	19.34
k_{cat}/K_m (s ⁻¹ · mM ⁻¹)	49.9	36.7	14.86	115.17
D-Galactose				
Co-substrate	NAD ⁺	NADP ⁺	–	–
K_m (mM)	0.57	0.44	–	–
V_{max} (units/mg)	90	55	–	–
k_{cat} (s ⁻¹)	61	37	–	–
k_{cat}/K_m (s ⁻¹ · mM ⁻¹)	108	85	–	–
D-Xylose				
Co-substrate	NAD ⁺	NADP ⁺	–	–
K_m (mM)	0.25	0.18	–	–
V_{max} (units/mg)	90	71	–	–
k_{cat} (s ⁻¹)	61	47	–	–
k_{cat}/K_m (s ⁻¹ · mM ⁻¹)	245	261	–	–
L-Arabinose				
Co-substrate	NAD ⁺	NADP ⁺	–	–
K_m (mM)	–	0.5	–	–
V_{max} (units/mg)	–	62	–	–
k_{cat} (s ⁻¹)	–	41	–	–
k_{cat}/K_m (s ⁻¹ · mM ⁻¹)	–	82	–	–

as well as the nucleotide-binding site (GXGXXG fingerprint motif; GDH-1 residues 188–193, GTGPIG) [13]. In GDH-2, Thr¹⁸⁹ is exchanged for serine, which has similar chemical properties [9]. In addition, GDH-2 harbours both of the residues (Asp²¹¹ and Arg²¹³ in GDH-1) that are discussed with respect to dual co-substrate specificity in GDH-1 by stabilizing the binding of the 2-phosphate of NADP⁺ [14]. Finally, the catalytic zinc-binding residues (Cys³⁹ and His⁶⁶ in GDH-1) are also found in both GDH isoenzymes.

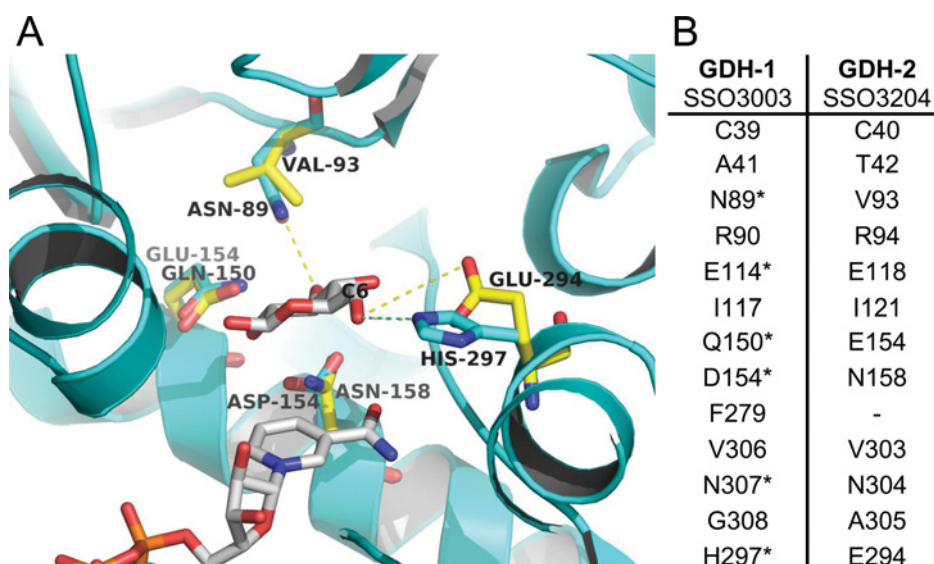
To address the molecular basis of why GDH-1 catalyses the oxidation of a range of different C₅ and C₆ sugars and GDH-2 only accepts glucose as substrate, a bioinformatic modelling approach was chosen. The three-dimensional structure of SSO3204 was modelled using the following procedure. First, the amino acid sequence was submitted to mGenTHREADER [18] to find possible structural templates. The highest scoring templates identified were the crystal structures of the GDH SSO3003 [14] (PDB codes 2CDB and 2CDC). Amino acid sequences of SSO3204 and SSO3003 were aligned [19], and the structure of SSO3204 was modelled based on this alignment and the structures of SSO3003 as template [20]. The model was visualized with PyMOL (DeLano Scientific; <http://www.pymol.org>). To test the robustness of the model, the same procedure was performed using the structure of GDH of the extreme

halophile *Haloferax mediterranei* [21] as template, which had the second best score in the mGenTHREADER search. Apart from small atomic displacements, this model was identical with the first one. The comparison of the vicinity of the sugar substrate in the crystal structure of GDH-1 and in the molecular model of GDH-2 shows a number of differences between both enzymes, including residues at positions that have been identified by Milburn et al. [14] as critical for the interaction with the sugar (Figure 2). Below, we discuss these differences and their likely effects on binding of glucose and xylose substrates.

The first important change is from Asn⁸⁹ in GDH-1 to Val⁹³ in GDH-2. This will probably result in a loss of a hydrogen bond between the sugar ring (C3-hydroxy group) and the enzyme and will thus weaken the affinity to β -glucose (pyranose) and to α/β -xylose (pyranose). Secondly, changes from Gln¹⁵⁰ to Glu¹⁵⁴, and from Asp¹⁵⁴ to Asn¹⁵⁸ were observed. These changes could be rather neutral with respect to substrate binding as, in both cases, the residues are exchanged for similar ones and the total charge is not affected. A more drastic change is that from the neutral or basic His²⁹⁷ in GDH-1 to the more acidic Glu²⁹⁴ in GDH-2, both contributed by a neighbouring polypeptide chain of the respective tetrameric enzymes. Whereas the histidine nitrogen ϵ 2 is involved in a hydrogen bond to the C6-hydroxy

Figure 2 | Comparison of sugar-binding sites

(A) Superposition of the glucose (centre, grey and red) -binding pocket of the crystal structure of GDH-1 (SSO3003, cyan) and model of the pocket in GDH-2 (SSO3204, yellow). (B) Amino acids within 0.5 nm of glucose in the crystal structure GDH-1 (PDB code 2CDB) and the amino acids at corresponding positions in the model of GDH-2. Asterisks mark key interactions with glucose according to Milburn et al. [14]. The residues in the last row are provided by a neighbouring polypeptide chain of the tetramer.



group of the glucose, it is plausible that the longer and more acidic glutamate group could lead to a stronger hydrogen bond to the same hydroxy group. This could compensate for the weaker interaction with the pyranose ring of glucose due to the transition of Asn⁸⁹ to Val⁹³ discussed above, so that GDH-2 still binds glucose. Xylose, in contrast with glucose, has no C6-hydroxy group, and thus cannot compensate for the loss of the hydrogen bond to Asn⁸⁹. The affinity of xylose to the enzyme should therefore be much reduced from GDH-1 to GDH-2. This is consistent with the observation that both enzymes use glucose as substrate, whereas only GDH-1 accepts xylose.

Physiological implications

Both isoenzymes, GDH-1 and GDH-2, catalyse the NAD⁺/NADP⁺-dependent oxidation of D-glucose. According to their enzymatic properties, GDH-2 seems to play the major role in D-glucose catabolism via the branched ED pathway, although GDH-1 might acquire important function at higher D-glucose concentrations and in the presence of NAD⁺. However, whereas GDH-2 is restricted to D-glucose degradation, GDH-1 seems to have additional functions in the catabolism of D-galactose, D-xylose and L-arabinose [9,14,22,23].

D-Galactose is degraded in *S. solfataricus* via the promiscuous ED pathway, whereas the metabolism of D-xylose is still unclear. In *Haloferax volcanii*, D-xylose

is degraded to 2-oxoglutarate using the Weimberg pathway (via xylose dehydrogenase, xylonolactonase, xylonate dehydratase, 2-oxo-3-deoxyxylonate dehydratase and 2-oxoglutarate-semialdehyde dehydrogenase) and therefore enters directly the citric acid cycle [24]. A homologue of xylose dehydrogenase is present in the genome of *S. solfataricus* (SSO3015); however, the respective activity is not confirmed. In addition, a similar pathway has been reported for D-arabinose degradation in this organism, and the arabinose-1-dehydrogenase has been characterized recently (SSO1300) [22]. The enzyme exhibits high affinity for D-arabinose ($K_m = 0.04$, $k_{cat}/K_m = 595 \text{ s}^{-1} \cdot \text{mM}^{-1}$), although also other sugars, i.e. L-fucose (6-desoxy-L-galactose), D-ribose and L-galactose are oxidized. Among others, arabinose-1-dehydrogenase was significantly up-regulated at gene (3.6-fold) and at protein (>20-fold) level in the presence of D-arabinose compared with D-glucose [22].

A new pathway for D-xylose and L-arabinose degradation has been reported recently [23]. The pathway has been reported to proceed via the multifunctional GDH-1, a substrate-specific dehydratase [D-xylonate dehydratase (SSO2665)] and aldol cleavage of 2-oxo-3-deoxypentionate to pyruvate and glycolaldehyde via KD(P)G aldolase, which, like GDH-1, exhibits broad substrate specificity [23]. Glycolaldehyde is converted into glyconate, glyoxylate and finally malate, which enters the citric acid cycle. Further on, *in vivo* NMR studies in *Sulfolobus acidocaldarius* revealed that D-xylose is degraded in parallel via the Weimberg

pathway, which omits aldol cleavage (see above and [24]). For L-arabinose degradation, the pathway (i.e. the specific L-arabinonate dehydratase) still has to be confirmed.

However, owing to the presence of several uncharacterized members of the pyridine-nucleotide-dependent alcohol/polyol/sugar dehydrogenase family in *S. solfataricus*, the role of GDH-1 in L-arabinose as well as D-xylose catabolism remains to be elucidated. It is tempting to speculate that the multifunctional GDH-1 might function as a 'jack of all trades', which is involved in different sugar degradation pathways in *S. solfataricus*, maybe as 'stand-by enzyme' in addition to pathway-specific sugar dehydrogenases, and allows for a quick adaptation to different carbon sources and thus fine-tuning of metabolism. This hypothesis is supported by the newly discovered D-glucose-specific GDH-2.

Funding

P.H., H.H. and M.Z. were supported by the Federal Ministry of Education and Research (BMBF) Germany in the course of the SulfoSYS project [SysMo P-N-01-09-23].

References

- Zillig, W., Stetter, K.O., Wunderl, S., Schulz, W., Priess, H. and Scholz, I. (1980) The *Sulfolobus*-*Caldariella* group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Arch. Microbiol.* **125**, 259–269
- She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M.J., Chan-Weiher, C.C.Y., Clausen, I.G., Curtis, B.A., De Moors, A. et al. (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7835–7840
- Wagner, M., Berkner, S., Ajon, M., Driessen, A.J.M., Lipps, G. and Albers, S.V. (2009) Expanding and understanding the genetic toolbox of the hyperthermophilic genus *Sulfolobus*. *Biochem. Soc. Trans.* **37**, 97–101
- Siebers, B. and Schönheit, P. (2005) Unusual pathways and enzymes of central carbohydrate metabolism in archaea. *Curr. Opin. Microbiol.* **8**, 695–705
- Zaparty, M., Esser, D., Gertig, S., Haferkamp, P., Kouril, T., Manica, A., Pham, T., Reimann, J., Schreiber, K., Sierocinski, P. et al. (2010) Hot standards for the thermoacidophilic archaeon *Sulfolobus solfataricus*. *Extremophiles* **14**, 119–142
- Snijders, A.P.L., Walther, J., Peter, S., Kinnman, I., de Vos, M.G.J., van de Werken, H.J.G., Brouns, S.J.J., Van Der Oost, J. and Wright, P.C. (2006) Reconstruction of central carbon metabolism in *Sulfolobus solfataricus* using a two-dimensional gel electrophoresis map, stable isotope labelling and DNA microarray analysis. *Proteomics* **6**, 1518–1529
- Albers, S.V., Birkeland, N.K., Driessen, A.J., Gertig, S., Haferkamp, P., Klenk, H.P., Kouril, T., Manica, A., Pham, T.K., Ruoff, P. et al. (2009) SulfoSYS (*Sulfolobus* systems biology): towards a silicon cell model for the central carbohydrate metabolism of the archaeon *Sulfolobus solfataricus* under temperature variation. *Biochem. Soc. Trans.* **37**, 58–64
- Ahmed, H., Ettema, T.J., Tjaden, B., Geerling, A.C., Van Der Oost, J. and Siebers, B. (2005) The semi-phosphorylative Entner–Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. *Biochem. J.* **390**, 529–540
- Lamble, H.J., Heyer, N.I., Bull, S.D., Hough, D.W. and Danson, M.J. (2003) Metabolic pathway promiscuity in the archaeon *Sulfolobus solfataricus* revealed by studies on glucose dehydrogenase and 2-keto-3-deoxygluconate aldolase. *J. Biol. Chem.* **278**, 34066–34072
- Lamble, H., Theodossis, A., Milburn, C., Taylor, G., Bull, S., Hough, D. and Danson, M. (2005) Promiscuity in the part-phosphorylative Entner–Doudoroff pathway of the archaeon *Sulfolobus solfataricus*. *FEBS Lett.* **579**, 6865–6869
- Kim, S. and Lee, S.B. (2005) Identification and characterization of *Sulfolobus solfataricus* D-gluconate dehydratase: a key enzyme in the non-phosphorylated Entner–Doudoroff pathway. *Biochem. J.* **387**, 271–280
- Kim, S. and Lee, S.B. (2006) Characterization of *Sulfolobus solfataricus* 2-keto-3-deoxy-D-gluconate kinase in the modified Entner–Doudoroff pathway. *Biosci. Biotechnol. Biochem.* **70**, 1308–1316
- Giardina, P., de Biasi, M.G., de Rosa, M., Gambacorta, A. and Buonocore, V. (1986) Glucose dehydrogenase from the thermoacidophilic archaeobacterium *Sulfolobus solfataricus*. *Biochem. J.* **239**, 517–522
- Milburn, C.C., Lambale, H.J., Theodossis, A., Bull, S.D., Hough, D.W., Danson, M.J. and Taylor, G.L. (2006) The structural basis of substrate promiscuity in glucose dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J. Biol. Chem.* **281**, 14796–14804
- Riveros-Rosas, H., Julian-Sanchez, A., Villalobos-Molina, R., Pardo, J.P. and Pina, E. (2003) Diversity, taxonomy and evolution of medium-chain dehydrogenase reductase superfamily. *FEBS J.* **270**, 3309–3334
- Chong, P.K., Burja, A.M., Radianingtyas, H., Fazeli, A. and Wright, P.C. (2007) Translational and transcriptional analysis of *Sulfolobus solfataricus* P2 to provide insights into alcohol and ketone utilization. *Proteomics* **7**, 424–435
- Kutschki, S. (2010) Purification and Characterization of a Glucose Dehydrogenase Isoenzyme of the Thermoacidophilic Crenarchaeon *Sulfolobus solfataricus*. Bachelor Thesis, University of Duisberg-Essen, Essen, Germany.
- McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) The psipred protein structure prediction server. *Bioinformatics* **16**, 404–405
- Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–17
- Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U. and Sali, A. (2006) Comparative protein structure modeling using Modeller. *Curr. Protoc., Bioinformatics Chapter 5*, Unit 5.6
- Britton, K.L., Baker, P.J., Fisher, M., Ruzhickov, S., Gilmour, D.J., Bonete, M.J., Ferrer, J., Pire, C., Esclapez, J. and Rice, D.W. (2006) Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile *Haloferax mediterranei*. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4846–51
- Brouns, S.J.J., Walther, J., Snijders, A.P.L., van de Werken, H.J.G., Willemens, H.L.D.M., Worm, P., de Vos, M.G.J., Andersson, A., Lundgren, M., Mazon, H.F.M. et al. (2006) Identification of the missing links in prokaryotic pentose oxidation pathways. *J. Biol. Chem.* **281**, 27378–27388
- Nunn, C.E.M., Johnsen, U., Schönheit, P., Fuhrer, T., Sauer, U., Hough, D.W. and Danson, M.J. (2010) Metabolism of pentose sugars in the hyperthermophilic archaea *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius*. *J. Biol. Chem.* **285**, 33701–33709
- Johnsen, U., Dambeck, M., Zaiss, H., Fuhrer, T., Soppa, J., Sauer, U. and Schönheit, P. (2009) D-Xylose degradation pathway in the halophilic archaeon *Haloferax volcanii*. *J. Biol.* **284**, 272–303

Received 15 August 2010
doi:10.1042/BST0390077

Danksagung

Abschließend möchte ich allen Personen danken, die direkt oder indirekt an dieser Arbeit beteiligt waren.

Ich danke allen derzeitigen und früheren Kollegen der **Arbeitsgruppe Bioinformatik** der Universität Duisburg-Essen, vor allem **Manuel Prinz** und **Ludwig Ohl** für ihre Denkanstöße und technische Hilfe, außerdem **Oliver Kuhn**, **Stefan Franke**, **Mike Nemec** für viele hilfreiche Diskussionen.

Ich danke allen Kooperationspartnern für die Bereitstellung der Daten und für die fruchtbare Zusammenarbeit.

Mein Dank gilt auch den Mitarbeitern der genutzten Großrechner LiDO und CRAY, allen voran **René Vergohsen** für die administrative Betreuung.

Ich danke, dass mir die Möglichkeit gegeben wurde, diese Arbeit durchzuführen und für die Betreuung durch **Daniel Hoffmann**. Er gab mir zusätzlich Freiraum für meine persönliche und familiäre Entwicklung.

Nicht zuletzt danke ich meiner gesamten **Familie** für jedwede Unterstützung.

Erklärungen

Erklärung:

Hiermit erkläre ich, gem. § 6 Abs. (2) f) der Promotionsordnung der Fakultäten für Biologie und Geografie, Chemie und Mathematik zur Erlangung des Dr. rer. nat., dass ich das Arbeitsgebiet, dem das Thema *Computergestützte Modellierung von Protein-Ligand-Systemen* zuzuordnen ist, in Forschung und Lehre vertrete und den Antrag von Karsten Sewczyk befürworte und die Betreuung auch im Falle eines Weggangs, wenn nicht wichtige Gründe dem entgegenstehen, weiterführen werde.

Essen, den _____

Unterschrift eines Mitgliedes der Universität Duisburg-Essen

Erklärung:

Hiermit erkläre ich, gem. § 7 Abs. (2) c) + e) der Promotionsordnung der Fakultäten für Biologie und Geografie, Chemie und Mathematik zur Erlangung des Dr. rer. nat., dass ich die vorliegende Dissertation selbständig verfasst und mich keiner anderen als der angegebenen Hilfsmittel bedient habe.

Essen, den _____

Unterschrift des Doktoranden

Erklärung:

Hiermit erkläre ich, gem. § 7 Abs. (2) d) + f) der Promotionsordnung der Fakultäten für Biologie und Geografie, Chemie und Mathematik zur Erlangung des Dr. rer. nat., dass ich keine anderen Promotionen bzw. Promotionsversuche in der Vergangenheit durchgeführt habe und dass diese Arbeit von keiner anderen Fakultät/Fachbereich abgelehnt worden ist.

Essen, den _____

Unterschrift des Doktoranden